

09/965594

STIC-Biotech/ChemLib

From: Schnizer, Richard
Sent: Wednesday, August 27, 2003 2:28 PM
To: STIC-Biotech/ChemLib
Subject: 09/965,594

Please search the commercial and published application databases for polypeptide SEQ ID NOS: 1, 12, 14, 16, 18, 20, 22, and 26 from 09/965,594. Please also search for nucleic acids that could encode these polypeptides.

Thank you-

Richard Schnizer, Ph.D.
Patent Examiner
Art Unit 1635
CM1 12E17
703-306-5441
Mail Box CM1 11E12

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: _____
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 30, 2003, 19:18:33 : Search time 2365.6 Seconds
(without alignments)
3147.423 Million cell updates/sec

Title: US-09-965-594-1

Perfect score: 953

Sequence: 1 MAPITAYAQOTRGLGCIIT.....GVAKAVDFIPVLSLETIMRS 182

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xl
-Q/cgn2.1/uspto_spo1/US09955594/runat_29082003_151919_28310/app_query.fasta.i.2872
-DB=GenEmbl -OPMT=fastap -SUPFLX=rge -MINMATCH=0.1 -LOOPECL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-LOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MFN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09955594@cgn.1.1.14686@runat_29082003_151919_28310 -NCPU=6 -ICPU=3
-NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -YGAPOP=6
-YGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:

1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.ua:*
14: gb.vi:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.mu:*
20: em.om:*
21: em.or:*
22: em.ov:*
23: em.pat:*
24: em.ph:*
25: em.pl:*
26: em.ro:*
27: em.sts:*
28: em.un:*

29: em.vi:*
30: em.htg_hum:*
31: em.htg_inv:*
32: em.htg_other:*
33: em.htg_mus:*
34: em.htg_pln:*
35: em.htg_rod:*
36: em.htg_mam:*
37: em.htg_vrt:*
38: em.sy:*
39: em.htgo_hum:*
40: em.htgo_mus:*
41: em.htgo_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	946	99.3	2058	6	AX395309 Sequence
2	946	99.3	2058	6	AX454818 Sequence
3	945	99.2	543	14	AF369218 Hepatitis
4	945	99.2	543	14	AF369235 Hepatitis
5	943	99.0	5360	6	AR118686 Sequence
6	943	99.0	5360	6	I06434 Sequence 48
7	943	99.0	5360	6	I09328 Sequence 8
8	943	99.0	6785	6	AR118692 Sequence
9	943	99.0	6785	6	I06440 Sequence 54
10	943	99.0	6785	6	I09329 Sequence 10
11	943	99.0	7310	6	AR118696 Sequence
12	943	99.0	7310	6	I09331 Sequence 15
13	943	99.0	7310	14	HPCPOLYP M32084 Hepatitis C
14	943	99.0	8316	6	AR118703 Sequence
15	943	99.0	8987	6	AR118728 Sequence
16	943	99.0	9185	6	AR118722 Sequence
17	943	99.0	9185	6	AR118723 Sequence
18	943	99.0	9185	6	BD091382 HCV culti
19	943	99.0	9185	6	I08294 Sequence 1
20	943	99.0	9379	6	AR166930 Sequence
21	943	99.0	9379	6	AR301300 Sequence
22	943	99.0	9401	6	AR176483 Sequence
23	943	99.0	9401	6	BD080334 Hepatitis
24	943	99.0	9401	6	E66593 Hepatitis C
25	943	99.0	9401	6	I71894 Sequence 9
26	943	99.0	9401	6	I81885 Sequence 9
27	943	99.0	9401	14	HPCPLYPRE M62321 Hepatitis C
28	943	99.0	9609	12	AF387805 Synthetic
29	943	99.0	9609	12	AF387808 Synthetic
30	943	99.0	9618	14	AF271632 Hepatitis
31	943	99.0	9646	12	AF387806 Synthetic
32	943	99.0	9693	12	AF387807 Synthetic
33	942	98.8	543	14	AF369222 Hepatitis
34	942	98.8	543	14	AF369232 Hepatitis
35	942	98.8	543	14	AF369240 Hepatitis
36	942	98.8	543	14	AF369245 Hepatitis
37	942	98.8	2061	6	AX441176 Sequence
38	942	98.8	2061	6	AX467113 Sequence
39	940	98.6	543	14	AF369224 Hepatitis
40	940	98.6	543	14	AF369237 Hepatitis
41	940	98.6	9424	14	AF511948 Hepatitis
42	939	98.5	543	14	AF369230 Hepatitis
43	939	98.5	9502	6	E08263 grNA of Hep
44	939	98.5	9502	6	E08264 CDNA of Hep
45	939	98.5	9502	14	HPCRCUJ D10749 Hepatitis C

ALIGNMENTS

RESULT 1


```

AX395309
LOCUS AX395309 2058 bp DNA linear PAT 18-MAY-2002
DEFINITION Sequence 2 from Patent WO0196875.
ACCESSION AX395309
VERSION AX395309.1 GI:21066308
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE
1 Chien,D.Y., Arcangel,P., Tandeske,L., George-Nascimento,C.,
Colt,D. and Medina-Selby,A.
Hcv antigen/antibody combination assay
Patent: WO 0196875-A 2 20-DEC-2001;
CHIRON CORPORATION (US)
FEATURES
source
1..2058
Location/Qualifiers
1..2058
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
/note="representative NS3/4a conformational antigen"
1..>2058
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/db_xref="GI:21066309"
/translation="MAPITAAQOTRGLGCIITSLTRDKNOVEGEVOIVSTAOTF
LATCINGVCWTVYHGACGTRTIASPKGVVOMYTNVDODLVGWPAQGSRLPTCTGGS
SDLYLVTRHADVIPIVRRGDSRGLISLSPRTISYKSGSGPLLCAGHAGVIFRAAIV
TRGVAKAVDFIPVENLETTMRSPVTDNSPPVPQSFQVAHLHAPTGGKSTKVPAA
YAAQGYKVLVLPNSVAATLFGFAYMSKAHGDIPNIRTVGRTITGSPITVYTKFLA
DGGCGGAYDIIICDECHSDATSLIGIGTVLQDAETAGARLVLAATAPPGSVTVPH
PNIIEVALSTTGEIPFYGKAIPLEVIKGGRLIFCHSKKCDLAKLVALGINAVAY
YRGLDVSVIPIGVDVVDALMTGVTGDFSDVDCNTCVTQTFDSDIPTFTIETI
YRGLDVSVIPIGVDVVDALMTGVTGDFSDVDCNTCVTQTFDSDIPTFTIETI
TLQDAVARTQRTGRTGKPGIYRFVAPGERPGSMFDSVCECCDAGCANWELTIPA
ETTVLRAYMNTPGVCCDHLFEWEGVFTGLTHIDAHFLSQTKQSGENLPYLVAOA
TVCAQAQPPPSWQDMKCLIRLKPTLHGPTPLYLRLGAVONRITLHPVTKYIMTCM
SADLEWVTSTWYLVGGVLAALAAAYCLSTGCVVIVGRVYLVSGKPAIIPDREVLYREFDE
MEEC"
BASE COUNT 419 a 634 c 580 g 425 t
ORIGIN

Alignment Scores:
Pred. No.: 6,56e-68 Length: 2058
Score: 946.00 Matches: 180
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 98.90% Mismatches: 0
Query Match: 99.27% Indels: 0
DB: 6 Gaps: 0

US-09-965-594-1 (1-182) x AX395309 (1-2058)

QY 1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
Db 1 ATGGGGCCCATACCGCGGTAGCCAGCAGACAGAGGGGCCCTCCTAGCGTGCATATACCC 60
QY 21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
Db 61 AGCCTAACTGGCGCGCAGCAAAACCAAGTGGAGGTGAGTCCAGATTGTCTCAACTGCT 120
QY 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
Db 121 GCCCAACACTTCTGGCAACGGTCATCATGCGGTGCTGGACTGCTACACGGGGCC 180
QY 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
Db 181 GGAACGAGGACCATCGCGTCACCAAGGCTCCTGTCCATCCAGATGATACCAATGTAGAC 240
QY 81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
Db 241 CAAGACCTTGTGGGTGGCGCGCTCCGGAAGGTAGCCGATCATTCACACCCCTGCACCTTC 300

```

```

QY 101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg 120
Db 301 GGCTCCTCGGACCTTTACCTGGTCACGAGCAGCGCATGTCATTCCTCGTGCCTCGCGCG 360
QY 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
Db 361 GGTATAGCAGGCGACCTGCTGCGCCCGGCCCATTTCTACTACTGAAGGCTCCTCG 420
QY 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
Db 421 GGGGTCCTGCTGTTGTCGCCCGCGGCGACGCGGTGGGCATATTTAGGCGCGCGGTGC 480
QY 161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
Db 481 ACCGCTGGAGTGGCTAAGCGGTGGACTTTATCCTGTGGAGAACCTAGACACCACTG 540
QY 181 ArgSer 182
Db 541 AGGTCC 546

RESULT 2
AX454818
LOCUS AX454818 2058 bp DNA linear PAT 06-JUL-2002
DEFINITION Sequence 1 from Patent WO0196870.
ACCESSION AX454818
VERSION AX454818.1 GI:21714047
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE
1 Chien,D.Y., Arcangel,P., Tandeske,L., George-Nascimento,C.,
Colt,D. and Medina-Selby,A.
Immunosays for anti-hcv antibodies
Patent: WO 0196870-A 1 20-DEC-2001;
CHIRON CORPORATION (US)
FEATURES
source
1..2058
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
/note="representative NS3/4a conformational antigen"
1..>2058
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAD38232.1"
/db_xref="GI:21714048"
/translation="MAPITAAQOTRGLGCIITSLTRDKNOVEGEVOIVSTAOTF
LATCINGVCWTVYHGACGTRTIASPKGVVOMYTNVDODLVGWPAQGSRLPTCTGGS
SDLYLVTRHADVIPIVRRGDSRGLISLSPRTISYKSGSGPLLCAGHAGVIFRAAIV
TRGVAKAVDFIPVENLETTMRSPVTDNSPPVPQSFQVAHLHAPTGGKSTKVPAA
YAAQGYKVLVLPNSVAATLFGFAYMSKAHGDIPNIRTVGRTITGSPITVYTKFLA
DGGCGGAYDIIICDECHSDATSLIGIGTVLQDAETAGARLVLAATAPPGSVTVPH
PNIIEVALSTTGEIPFYGKAIPLEVIKGGRLIFCHSKKCDLAKLVALGINAVAY
YRGLDVSVIPIGVDVVDALMTGVTGDFSDVDCNTCVTQTFDSDIPTFTIETI
TLQDAVARTQRTGRTGKPGIYRFVAPGERPGSMFDSVCECCDAGCANWELTIPA
ETTVLRAYMNTPGVCCDHLFEWEGVFTGLTHIDAHFLSQTKQSGENLPYLVAOA
TVCAQAQPPPSWQDMKCLIRLKPTLHGPTPLYLRLGAVONRITLHPVTKYIMTCM
SADLEWVTSTWYLVGGVLAALAAAYCLSTGCVVIVGRVYLVSGKPAIIPDREVLYREFDE
MEEC"
BASE COUNT 419 a 633 c 581 g 425 t
ORIGIN

Alignment Scores:
Pred. No.: 6,56e-68 Length: 2058
Score: 946.00 Matches: 180
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 98.90% Mismatches: 0
Query Match: 99.27% Indels: 0
DB: 6 Gaps: 0

US-09-965-594-1 (1-182) x AX454818 (1-2058)

```

```

Qy 1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
Db 1 ATGGCGCCATCAGCGGTACGCCACACAGAGGGCCCTCTAGGTGATATACACC 60
Qy 21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
Db 61 AGCCTAACTGGCGGGACAAACCAAGTGGAGGTGAGGTCCAGATTGTCTCAACTGCT 120
Qy 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
Db 121 GCCCAAAACCTTCCTGGCAACGTGCATCAATGGGGTGCTGGAGTGTACACAGGGGCC 180
Qy 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
Db 181 GGAAGGAGGACCATCGGTACCAAGGTCCTGTATCCAGATGTATACCAATGTAGAC 240
Qy 81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
Db 241 CAAGACCTTGTGGCTGGCGGCTCCGCAAGGTAGCCCATCATTGACACCTGCACCTTGC 300
Qy 101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg 120
Db 301 GGCTCCTCGGACCTTACCTGGTCACGAGGACCGCCGATGTATCCCGTCCCGCGCGG 360
Qy 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
Db 361 GGTGATAGCAGGGCAGCCTCTGTGCGCGCGGCCATTCCTACTTGAAGGCTCCTCG 420
Qy 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
Db 421 GGGGTCTGGCTGTGTGCGCGCGGCGGACCGCGGCGCATATTAGGGCGCGCGTGTGC 480
Qy 161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
Db 481 ACCCGTGGAGTGGCTAAGCGGTGGACTTATCCCTGTGGAGAACCTAGAGACCAACCTG 540
Qy 181 ArgSer 182
Db 541 AGGTCC 546

RESULT 3
AF369218 543 bp RNA linear VRL 03-JUL-2002
LOCUS Hepatitis C virus PL.4 NS3 protease gene, partial cds.
DEFINITION AF369218
ACCESSION AF369218
VERSION AF369218.1 GI:14150560
KEYWORDS Hepatitis C virus
SOURCE Hepatitis C virus
ORGANISM Hepatitis C virus
REFERENCE 1 (bases 1 to 543)
AUTHORS Holland-Staley,C.A., Kovari,L.C., Golenberg,E.M., Pobursky,K.J. and Mayers,D.L.
TITLE Genetic diversity and response to IFN of the NS3 protease gene from clinical strains of the hepatitis C virus
JOURNAL Arch. Virol. 147 (7), 1385-1406 (2002)
MEDLINE 22105140
PUBMED 1211114
REFERENCE 2 (bases 1 to 543)
AUTHORS Holland-Staley,C.A., Kovari,L.C., Golenberg,E. and Mayers,D.L.
TITLE Direct Submission
JOURNAL Submitted (09-APR-2001) Infectious Disease Research, Henry Ford Health Systems, 2799 W. Grand Blvd. Rm 7045 E & R, Detroit, MI 48202, USA
FEATURES
    Location/Qualifiers
        1..543
            /organism="Hepatitis C virus"
            /mol_type="genomic RNA"
            /isolate="pt.4"
            /db_xref="taxon:11103"
            /note="type: 1A"

```

```

<1..>543
/note="polyprotein"
/codon_start=1
/product="NS3 protease"
/protein_id="AAK54543.1"
/db_xref="GI:14150561"
/translation="APITAYAQQTGRGLGCLITSLTGRDKNQVEGVOIVSTAAQTFPL
ATCINGVCWTYVHGAGTFTIASPKGPVIQMTYNDKDLVGPAPQGSRSILTPCTCGSS
DLXYTRHADVIPVRECDSRGSLSPRPISYLAGSSGGPLLCPAGHAVGIFRAVCT
RGVAKAVDFIPVENLETHRS"
BASE COUNT 113 a 166 c 158 g 106 t
ORIGIN

Alignment Scores:
Pred. No.: 1..89e-68 Length: 543
Score: 945.00 Matches: 180
Percent Similarity: 100.00% Conservatives: 1
Rest Local Similarity: 99.45% Mismatches: 0
Query Match: 99.16% Indels: 0
DB: 14 Gaps: 0

US-09-965-594-1 (1-182) x AF369218 (1-543)
Qy 2 AlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThrSer 21
Db 1 GCACCTATCACGGCATACGCCACAGACAGAGGGCCCTCTAGGTGTATATACCACT 60
Qy 22 LeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAla 41
Db 61 CTAACTGGCGGGACAAACCAAGTGGAGGTGAGGTCCAGATTGTGTCACCTGCTGCC 120
Qy 42 GlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGly 61
Db 121 CAGACTTCTTGGCAACATGCATCAACGGGTATGCTGGACCGTCTACACGGGGCCGA 180
Qy 62 ThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAspLys 81
Db 181 ACAGAGACCATTCATCACCAAGTCCCGTCATCCAGATGTACCAATGTATACAAA 240
Qy 82 AspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCysGly 101
Db 241 GATCTGTGGGTGGCGCGCTCTCAAGGTTCCTCATTCACACCTGCACCTCGGC 300
Qy 102 SerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgGly 121
Db 301 TCCTCGGACCTTACTTGTGTGTCAGAGCATGCGATGTATACCGGTGCGCGCGAGGT 360
Qy 122 AspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerGly 141
Db 361 GATAGCAGGGGAGCGCTGTCTGCGCGCGGCCCATTTCTACTTGAAGGCTCTCGGG 420
Qy 142 GlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThr 161
Db 421 GGTCCGCTGTGTGCGCGCGGACACGCGGTAGGCATATTACGCGCGCGGTGTGCACC 480
Qy 162 ArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMetArg 181
Db 481 CGTGAGTGGCTAAGCGGGTGACCTTATCCCGTGGAGAACCTAGAGACCAACCACTAGG 540
Qy 182 Ser 182
Db 541 TCC 543

RESULT 4
AF369235 543 bp RNA linear VRL 03-JUL-2002
LOCUS Hepatitis C virus pt.1Y NS3 protease gene, partial cds.
DEFINITION AF369235
ACCESSION AF369235
VERSION AF369235.1 GI:14150594
KEYWORDS Hepatitis C virus
SOURCE Hepatitis C virus
ORGANISM Viruses: ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

```

Hepacivirus.
 1 (bases 1 to 543)
 Holland-Staley,C.A., Kovari,L.C., Golenberg,E.M., Pobursky,K.J. and
 Mayers,D.L.
 Genetic diversity and response to IFN of the NS3 protease gene from
 clinical strains of the hepatitis C virus
 Arch. Virol. 147 (7), 1385-1406 (2002)
 22105140
 MEDLINE
 12111414
 2 (bases 1 to 543)
 Holland-Staley,C.A., Kovari,L.C., Golenberg,E. and Mayers,D.L.
 Direct Submission
 Submitted (09-APR-2001) Infectious Disease Research, Henry Ford
 Health Systems, 2799 W. Grand Blvd. Rm 7045 E & R, Detroit, MI
 48202, USA
 FEATURES
 Location/Qualifiers
 source
 1..543
 /organism="Hepatitis C virus"
 /mol_type="genomic RNA"
 /isolate="pt. 1v"
 /db_xref="taxon:11103"
 /note="type: 1A"
 <1..>543
 /note="polyprotein"
 /codon_start=1
 /product="NS3 protease"
 /protein_id="AAK54560.1"
 /db_xref="GI:14150595"
 /translation="APITAYAAQTRGLLCCITSLTGRDXNQVEGEVOIVSTAAQTFL
 ATCINGCVMTVTHGAGTRTIAASPGKPVIOYTNVDKDLVGMFAPGSRSLTPTCCGSS
 DIYLYTRHADVTPVRRRGRSGSLSPRIYLYKSSGGPLLCFAGHAHVGFRAAVCT
 RGVAKAVDFIPVENLETTMRS"
 BASE COUNT 112 a 169 c 159 g 103 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,89e-68 Length: 543
 Score: 945.00 Matches: 180
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 99.45% Mismatches: 0
 Query Match: 99.16% Indels: 0
 DB: 14 Gaps: 0
 US-09-965-594-1 (1-182) x AF369235 (1-543)
 Qy 2 AlaProleThrAlaTyAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThrSer 21
 Db 1 GCACCATCACGATACATACCCAGCAGCAGCAAGGGGCTCTAGGGTGCAATACACCAGC 60
 Qy 22 LeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAla 41
 Db 61 CTACCGGGCGGGGACAAAACCAAGTGGAGGGTGAGGTCCAGATGTGTCAACTGCTGCC 120
 Qy 42 GlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValThrHisGlyAlaGly 61
 Db 121 CAACITTTCTTGCACAGTCATCAATGGGGTATGGTCCAGCCGTCTACCCAGCGCCGGA 180
 Qy 62 ThrArgThrIleAlaSerProLysGlyProValIleGlnMetThrThrAsnValAspLys 81
 Db 181 ACAGGACCATCGATCATCAACAAAGGTCTCCGTATCCAGATGTACACCAATGTAGACAAA 240
 Qy 82 AspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrThrProCysThrCysGly 101
 Db 241 GATCTTGGGGCTGGCCCGCTCTCAAGGTTCCCGCTCAITGACACCCCTGCACCTCGCGG 300
 Qy 102 SerSerAspLeuTyIleuValThrArgHisAlaAspValIleCProValArgArgGly 121
 Db 301 TCCTCGACCTTTACCTGGTGCAGAGCAGCGCGGATGATCCCGTGCSCCGGGGGT 360
 Qy 122 AspSerArgGlySerLeuLeuSerProArqProIleSerTyIleuLysGlySerGly 141
 Db 361 GATAGCAGGGGAGCCTGTCTCGCCCGGCCCATTTCTACTTGAAGGCTCCTCGGGG 420

Qy 142 GlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThr 161
 Db 421 GGTCCCTCTTTGGCCCGGGACACCGGTAGGCATATTACAGGCGCGGTGTGCACC 480
 Qy 162 ArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMetArg 181
 Db 481 CGTGGAGTGGCTAAGCGGTGGACITTTATCCCTCTGGAGAACCTAGACACCAACCATGAGG 540
 Qy 182 Ser 182
 Db 541 TCC 543
 RESULT 5
 AR118686
 LOCUS AR118686
 DEFINITION Sequence 53 from patent US 6150087.
 ACCESSION AR118686
 VERSION AR118686.1 GI:14100596
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 5360)
 AUTHORS Chien,D.Y.
 TITLE NANBV diagnostics and vaccines
 JOURNAL Patent: US 6150087-A 53 21-NOV-2000;
 FEATURES Location/Qualifiers
 source
 1..5360
 /organism="unknown"
 BASE COUNT 1060 a 1623 c 1532 g 1145 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 3,22e-67 Length: 5360
 Score: 943.00 Matches: 179
 Percent Similarity: 100.00% Conservative: 3
 Best Local Similarity: 98.35% Mismatches: 0
 Query Match: 98.95% Indels: 0
 DB: 6 Gaps: 0
 US-09-965-594-1 (1-182) x AR118686 (1-5360)
 Qy 1 MetaLapProIleThrAlaTyAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
 Db 930 CTGGCGCCCATCAGCGGTACGCCAGCAGCAAGGGGCTCTAGGGTGCAATATCACC 989
 Qy 21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
 Db 990 AGCTTAAGTGGCGGGACAAAACCAAGTGGAGGGTGGTCCAGATTGTGTCAACTGCT 1049
 Qy 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValThrHisGlyAla 60
 Db 1050 GCCCAACCTCTCTGGCAAGTCATCAATGGGGTGTGTGGACTGTCTACCCAGGGGCC 1109
 Qy 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetThrThrAsnValAsp 80
 Db 1110 GGAACAGAGGACATCGCGTCCACCAAGGGTCTGTGTATCCAGATGTATACCAATGTAGAC 1169
 Qy 81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
 Db 1170 CAAGACCTTGTGGCTGGCCCGCTCCGCAAGGTAGCCGCTCATTTGACACCCCTGCCTTGC 1229
 Qy 101 GlySerSerAspLeuTyIleuValThrArgHisAlaAspValIleProValArgArgArg 120
 Db 1230 GGCTCTCTGGAGCCTTTACCTGGTGCAGGACGCGCGATGTATCCCTGCGCGGGCGG 1289
 Qy 121 GlyAspSerArgGlySerLeuLeuSerProArqProIleSerTyIleuLysGlySerSer 140
 Db 1290 GGTGATAGAGGGGACGCTGTGTGCGCCCGGCCCATTTCTACTTGAAGGCTCCTCGG 1349
 Qy 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
 Db 1350 GGGGGTCCCTGTGTGTCGCCCGGGGACCGCGTGGCATATTTAGGCGCGCGGTGTGC 1409

Db 1470 AGGTCC 1475

RESULT 8
AR118692
LOCUS AR118692 6785 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 65 from patent US 6150087.
ACCESSION AR118692
VERSION AR118692.1 GI:14100602
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 6785)
AUTHORS Chien D.Y.
TITLE NABV diagnostics and vaccines
JOURNAL Patent: US 6150087-A 65 21-NOV-2000;
FEATURES
Location/Qualifiers
1..6785
/organism="unknown"
BASE COUNT 1392 a 2050 c 1914 g 1429 t
ORIGIN

Alignment Scores:
Pred. No.: 4.15e-67 Length: 6785
Score: 943.00 Matches: 179
Percent Similarity: 100.00% Conservative: 3
Best Local Similarity: 98.35% Mismatches: 0
Query Match: 98.95% Indels: 0
DB: 6 Gaps: 0

US-09-965-594-1 (1-182) x AR118692 (1-6785)

QY 1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
:::|||||
Db 1203 CTGGCGCCCATCACGGGTACGCCAGCAGACAAAGGGGCTCTAGGGTCATAATCACC 1262

QY 21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
:::|||||
Db 1263 AGCCTAACTGGCGGGACAAACCAAGTGGAGGTGAGGTCCAGATGTGTCAACTGCT 1322

QY 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
:::|||||
Db 1323 GCCCAACCTTCTTGGCAACGTCATCAATGGGTGTGCTGAGCTGTCTACACGGGGCC 1382

QY 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
:::|||||
Db 1383 GGAACGAGGACCATCGGCTCACCAAGGGTCTCTCATCCAGATGTATACCAATGTAGAC 1442

QY 81 LysAspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
:::|||||
Db 1443 CAAGACCTTGTGGCTGGCGCTCCGCAAGGTAGCGGCTCATTCACACCGCTGCACCTGC 1502

QY 101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 120
:::|||||
Db 1503 GGCTCTCGGACCTTTACTGTCTCAGGAGCAGCCGCGATGTCTATCCGCGCGCGCGG 1562

QY 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
:::|||||
Db 1563 GGTGATAGCAGGGCAGCTGTCTGCGCGCGCGCGGCAAGTCTCTACTTGAAGGCTCTCG 1622

QY 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
:::|||||
Db 1623 GGGGTCTGGCTGTGTGCGCGCGCGGCGCACCGCTGGGCATATTAGGGCGCGGTGTC 1682

QY 161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
:::|||||
Db 1683 ACCCTGGAGTGGCTAAGCGGGTGGACTTTATCCCTGTGGAGAACCTTAGACACCATG 1742

QY 181 ArgSer 182
|||||
Db 1743 AGGTCC 1748

RESULT 10
109329
LOCUS 109329 6785 bp DNA linear PAT 02-DEC-1994
DEFINITION Sequence 10 from patent WO 8904669.

Db 1470 AGGTCC 1475

RESULT 8
AR118692
LOCUS AR118692 6785 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 65 from patent US 6150087.
ACCESSION AR118692
VERSION AR118692.1 GI:14100602
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 6785)
AUTHORS Chien D.Y.
TITLE NABV diagnostics and vaccines
JOURNAL Patent: US 6150087-A 65 21-NOV-2000;
FEATURES
Location/Qualifiers
1..6785
/organism="unknown"
BASE COUNT 1392 a 2050 c 1914 g 1429 t
ORIGIN

Alignment Scores:
Pred. No.: 4.15e-67 Length: 6785
Score: 943.00 Matches: 179
Percent Similarity: 100.00% Conservative: 3
Best Local Similarity: 98.35% Mismatches: 0
Query Match: 98.95% Indels: 0
DB: 6 Gaps: 0

US-09-965-594-1 (1-182) x AR118692 (1-6785)

QY 1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
:::|||||
Db 1203 CTGGCGCCCATCACGGGTACGCCAGCAGACAAAGGGGCTCTAGGGTCATAATCACC 1262

QY 21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
:::|||||
Db 1263 AGCCTAACTGGCGGGACAAACCAAGTGGAGGTGAGGTCCAGATGTGTCAACTGCT 1322

QY 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
:::|||||
Db 1323 GCCCAACCTTCTTGGCAACGTCATCAATGGGTGTGCTGAGCTGTCTACACGGGGCC 1382

QY 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
:::|||||
Db 1383 GGAACGAGGACCATCGGCTCACCAAGGGTCTCTCATCCAGATGTATACCAATGTAGAC 1442

QY 81 LysAspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
:::|||||
Db 1443 CAAGACCTTGTGGCTGGCGCTCCGCAAGGTAGCGGCTCATTCACACCGCTGCACCTGC 1502

QY 101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 120
:::|||||
Db 1503 GGCTCTCGGACCTTTACTGTCTCAGGAGCAGCCGCGATGTCTATCCGCGCGCGCGG 1562

QY 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
:::|||||
Db 1563 GGTGATAGCAGGGCAGCTGTCTGCGCGCGCGGCGCACCGCTGGGCATATTAGGGCGCGGTGTC 1622

QY 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
:::|||||
Db 1623 GGGGTCTGGCTGTGTGCGCGCGCGGCGCACCGCTGGGCATATTAGGGCGCGGTGTC 1682

QY 161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
:::|||||
Db 1683 ACCCTGGAGTGGCTAAGCGGGTGGACTTTATCCCTGTGGAGAACCTTAGACACCATG 1742

QY 181 ArgSer 182
|||||
Db 1743 AGGTCC 1748

RESULT 9

```

ACCESSION      I09329
VERSION        I09329.1  GI:587964
KEYWORDS
SOURCE        Unknown.
ORGANISM      Unknown.
REFERENCE      1 (bases 1 to 6785)
AUTHORS      Houghton, M., Choo, Q.-K. and Kuo, G.
JOURNAL      Patent: WO 8904669-A 10 01-JUN-1989;
FEATURES      Location/Qualifiers
               source
               1..6785
               /organism="unknown"
BASE COUNT    1392 a 2050 c 1914 g 1429 t
ORIGIN
Alignment Scores:
Pred. No.:    4,150-67      Length:    6785
Score:        943.00      Matches:    179
Percent Similarity: 100.00%  Conservative: 3
Best Local Similarity: 98.35%  Mismatches: 3
Query Match:   98.95%      Indels:    0
DB:           6           Gaps:      0

US-09-965-594-1 (1-182) x I09329 (1-6785)
Qy 1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
Db 1203 CTGGCGCCCATCAGCGCGTACGCCAGCAGACAGGGGCCCTCTAGGGTGCATATCACC 1262
Qy 21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
Db 1263 AGCTTAACCTGGCGGGGACAAAACCAAGTGGAGGTGAGGTCCAGATTGTGTCAACTGCT 1322
Qy 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValThrHisGlyAla 60
Db 1323 GCCCAAAACCTTCTCGCAACGTGCATCAATGGGTGTGCTGACACTGTCTACCCAGGGGCC 1382
Qy 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetIleThrAsnValAsp 80
Db 1383 GGAACGAGGACCATCGCTCACCAAGGCTCTGTCATCCAGATGTATACCAATGTAGAC 1442
Qy 81 LysAspLeuValGlyTrpProIleProGlnGlySerArgSerLeuThrProCysThrCys 100
Db 1443 CAAGACCTTGTGGGCTGGCCGCTCCGCAAGGTAGCCGCTCATTCACACCTTGCACTTGC 1502
Qy 101 GlySerSerAspLeuTyrIleValThrArgHisAlaAspValIleProValArgArg 120
Db 1503 GGCTCTCGGACCTTTACCTGTGTCAGAGGACGCGGATGTATTCCCGTGGCGGGCGG 1562
Qy 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrIleuLysGlySerSer 140
Db 1563 GGTGATAGCAGGGGACGCTGTGTCGCCCGGCCCAATTTCTTACTTGAAGGCTCCTCG 1622
Qy 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
Db 1623 GGGGGTCCGCTGTGTGCCCGCGGGCAGCGCGTGGGCATATTTAGGGCCGCGGTGTC 1682
Qy 161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
Db 1683 ACCCGTGGACTGGCTAAGCGGTGGACTTATATCCCTGTGGAGAACCTAGAGACAACCATG 1742
Qy 181 ArgSer 182
Db 1743 AGGTCC 1748

RESULT 11
AR118696
LOCUS        AR118696      7310 bp      DNA      linear      PAT 16-MAY-2001
DEFINITION   Sequence 74 from patent US 6150087.
ACCESSION   AR118696
VERSION     AR118696.1  GI:14100606
KEYWORDS
SOURCE      Unknown.

```

```

ORGANISM      Unknown.
REFERENCE      1 (bases 1 to 7310)
AUTHORS      Chien, D.Y.
TITLE        NANBV diagnostics and vaccines
JOURNAL      Patent: US 6150087-A 74 21-NOV-2000;
FEATURES      Location/Qualifiers
               source
               1..7310
               /organism="unknown"
BASE COUNT    1495 a 2220 c 2056 g 1539 t
ORIGIN
Alignment Scores:
Pred. No.:    4,5e-67      Length:    7310
Score:        943.00      Matches:    179
Percent Similarity: 100.00%  Conservative: 3
Best Local Similarity: 98.35%  Mismatches: 0
Query Match:   98.95%      Indels:    0
DB:           6           Gaps:      0

US-09-965-594-1 (1-182) x AR118696 (1-7310)
Qy 1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
Db 1728 CTGGCGCCCATCAGCGGTACGCCAGCAGACAGGGGCCCTCTAGGGTGCATATCACC 1787
Qy 21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
Db 1788 AGCTTAACCTGGCGGGGACAAAACCAAGTGGAGGTGAGGTCCAGATTGTGTCAACTGCT 1847
Qy 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValThrHisGlyAla 60
Db 1848 GCCCAAAACCTTCTCGCAACGTGCATCAATGGGTGTGCTGACACTGTCTACCCAGGGGCC 1907
Qy 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetIleThrAsnValAsp 80
Db 1908 GGAACGAGGACCATCGCTCACCAAGGCTCTGTCATCCAGATGTATACCAATGTAGAC 1967
Qy 81 LysAspLeuValGlyTrpProIleProGlnGlySerArgSerLeuThrProCysThrCys 100
Db 1968 CAAGACCTTGTGGGCTGGCCGCTCCGCAAGGTAGCCGCTCATTCACACCTTGCACTTGC 2027
Qy 101 GlySerSerAspLeuTyrIleValThrArgHisAlaAspValIleProValArgArg 120
Db 2028 GGCTCTCGGACCTTTACCTGTGTCAGAGGACGCGGATGTATTCCCGTGGCGGGCGG 2087
Qy 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrIleuLysGlySerSer 140
Db 2088 GGTGATAGCAGGGGACGCTGTGTCGCCCGGCCCAATTTCTTACTTGAAGGCTCCTCG 2147
Qy 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
Db 2148 GGGGGTCCGCTGTGTGCCCGCGGGCAGCGCGTGGGCATATTTAGGGCCGCGGTGTC 2207
Qy 161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
Db 2208 ACCCGTGGAGTGGCTAAGCGGTGGACTTATATCCCTGTGGAGAACCTAGAGACAACCATG 2267
Qy 181 ArgSer 182
Db 2268 AGGTCC 2273

RESULT 12
I09331
LOCUS        I09331      7310 bp      DNA      linear      PAT 02-DEC-1994
DEFINITION   Sequence 15 from Patent WO 8904669.
ACCESSION   I09331
VERSION     I09331.1  GI:587966
KEYWORDS
SOURCE      Unknown.
ORGANISM    Unknown.
REFERENCE   1 (bases 1 to 7310)

```

AUTHORS Houghton, M., Choo, Q.-K. and Kuo, G.
JOURNAL Patent: WO 8904669-A 15 01-JUN-1989;

FEATURES
Source Location/Qualifiers
1..7310
/organism="unknown"
BASE COUNT 1495 a 2218 c 2058 g 1539 t
ORIGIN

Alignment Scores:
Pred. No.: 4,5e-67 Length: 7310
Score: 943.00 Matches: 179
Percent Similarity: 100.00% Conservativity: 3
Best Local Similarity: 98.35% Mismatches: 0
Query Match: 98.95% Indels: 0
DB: 6 Gaps: 0

US-09-965-594-1 (1-182) x 109331 (1-7310)

QY 1 MetaAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
Db 1728 CTGGCGCCCATCACGGCTACGCCAGCAGACAGGGCCCTCTAGGTCATCAATACACC 1787
QY 21 SerLeuThrGlyArgAspValAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
Db 1788 AGCCTACTGGCGGAGCAAAACCAAGTGGAGGGTGAGGTCCAGATTGTCAACTGCT 1847
QY 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
Db 1848 GCCCAACCTTCTGGCAACGTGCATCAATCGGTGTCTGGAGCTGTCTACCAAGGGCC 1907
QY 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
Db 1908 GGAACGAGGACCATCGCTCACCAAGGGTCTCTGATCCAGATGTATACCAATGTAGAC 1967
QY 81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
Db 1968 CAAGACCTTGTGGCTGGCGCCCTCCGACAGTAGAGCCGTCATTGACACCTGCACCTGC 2027
QY 101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg 120
Db 2028 GGCCTCTGGACCTTACTGTGTGGCGGCGGACCCGATGTCATTCGCCGCGCGCGCG 2087
QY 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
Db 2088 GTGTATACAGGGGAGCGCTCTGTGGCGCCCGCCCATTTCTACTTGAAGGCTCTCTCG 2147
QY 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
Db 2148 GGGGCTCGGCTGTGTGGCGGCGGCGGACCCGCTGGGCATATTAGGGCGCGGTGTC 2207
QY 161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
Db 2208 ACCCTGTAGTGGCTTAAGCGCGTGGACTTTA:CCCTGTGGAGAACCTAGACACCACTG 2267
QY 181 ArgSer 182
Db 2268 AGGTCC 2273

RESULT 13
HPCPOLYP
LOCUS Hepatitis C virus polyprotein gene, partial cds. VRL 02-AUG-1993
DEFINITION M32084
ACCESSION M32084.1 G1:329875
VERSION
KEYWORDS polyprotein.
SOURCE Hepatitis C virus
ORGANISM Hepatitis C virus
Viruses: ssRNA positive-strand viruses, no DNA stage. Flaviviridae;
Hepadnavirus.
Hepatitis C virus
1 (bases 1 to 7310)

REFERENCE
AUTHORS Choo, Q.-K., Richman, K. and Han, J.
TITLE The nucleotide sequence of the Hepatitis C viral genome
JOURNAL Unpublished (1990)

COMMENT

Original source text: Hepatitis C virus, cDNA to viral RNA, clones K9-1 through 15e, isolated from chimpanzee (individual 910) blood plasma.

Draft entry and printed sequence for [1] kindly submitted by M.Houghton, 22-FEB-1990. Chiron Corporation, 4560 Horton Street, Emeryville CA 94608.

FEATURES

source

1..7310
/organism="Hepatitis C virus"
/mol_type="genomic RNA"
/db_xref="taxon:11103"
<1..>7310
/note="polyprotein"
/codon_start=3
/protein_id="AAA45677.1"
/db_xref="GI:329876"
/translation="GCPERLASCKPLTDFQGMGPISYANGSGPDQPCWHYPPKPC
GIVPAKSCGVPYCTPDSVVTGTDGAPTSWGENDTDFVNLNTPPLGNNWFGC
TWNSTGTGKVCAGPCVIGGAGNTLHCTDFCRKHDPATYSCGSDPMTPLCLVD
YPRLMHYPTCTINVTIFKIMYGVGVEHLEAECNTRGECDELDRESELSPLLT
TDQVLPCTFTLPALSTGLIHLHONIVDQYLYGVGSSIASWAKWEEVVLFLLL
ADARVCSLMMHLISQEAALENLILNAAISLAGTGLVFLVFFCFAYILGKNWP
GAVTTFYGMPLLLLLALPORAYALDTEVAASCGVLYGLMALTSFYRYKYSWC
LWMLQYITRVEAQLHWIPPLNVRGDRVILIMCAVHPTLVFDITKLIVAFGLW
ILOASLLKPYFVVGGLRFLCARMKIGHYVQWYIKIGALTCTYVNLTPLRD
WAHNGLRDLAVAVPVSQMETKLTWAGDTAACGDIINGLPVSRARRGELIGPAD
GMVSGRLLAPITAYAQOTGLLGCITSLTGDKNOVEGEVQIVSTAAQTFLATCI
NGVQWYVHCAGTFTIAPKGPVLOMYNDQDLVGMPPAQSGRSLSLPTCCSSDYL
KVRADIVPVRGDSRGSLLSPRISYLAGSGGGLPCAGHAGVGFRAAAYRQVA
KATDFIPVENLETTMRSPVTDNSPPVPSQFVAHLHAPTGHGKSTFPAAYAAQV
YKVLVPSVAATLFGAYMSKAGIDNPRTGVTITTTGSPITSTYKGLFADGSCS
GGATYIICDECHSDATSLGIGTQDAETAGARLVLATTPGTSVTPHPNTEE
VALSTGTEIPYKAIPLEVIKGRHLIFCHSKKQDELAALKALVGINAVYRGLD
VSVPTSGVVVATDALMTGYTGDPSVIDCNTCTVTQVDFSLDPTFTIETLRQD
AVSTRRGRTGRKPGIYRFVAPGRPSCHMFDSSVLCQYDAGCAWYELTPTATVR
LRVYNTPGLPVCDHLEFEGVFTGLTHDAHFLSOTKSGENLPYLVAYQATVCAR
AQAAPPDQMMKCLIRLFTLHGFTPLDLRLGAYQNEITLTHPTKTIIMTMSADLE
VVTSTVLVGGVLAALAAAYCLTSCGVVIVGSLRKPALIPREVLRYEFDEMECS
QHLPTIYQGMMLAEQFKALGLLQATASQARVIAPAVQTNQKLETFWAKHWNFI
GIQYLAGSLTLPNPAIASLMAETAAVTSPLTTSQTLLENILGGWAAQALAPGAAT
FVAGLAGAAIGSVGLKGLVIDILAGYAGVAGALVAFKIMSGEVPSTBDLVNLLPAI
LSPCALVGVVCAAILRRHVGCGEAVQWMNRLIAFASGNHVSPTHYVPESDAARV
TALLSSLTQTLRLRHLQWISSECTTFCSSWLRDIDWICEVLSDFKTLKAKLMPQ
LPFGVSCQGYGKGVWVDGIMHRCCHGAEITGHVKNGTMRIVGPRCTRNNMSGTF
PINATYDGPCTPLPAPNTFALMRVSAEYVEIRQVDGFHYVTGMTDNLKPCQVPS
PEFTELDGVRLHRAFPCKPLREYVSFRVGHYEPVSQLCEPPEPDAVVTSLMT
DPSHITAEAGRRLARGSPSSASSASQSLKATCANHSDPSDAELIEANILWR
QMGNGNTRVESKNKVIWLSDFDLVAEEDEREISVPAEILRKRRAFOALPVWARP
YNPLVETKKPDYPPVHGCPLPPKPPSPVPPPKRTVLTSTLSTALAEALATR
SFGSSSTSGITGDNATTSSEPAEGCPCPDSESYSSMPLEGEPPDLSGDSNSTV
SSEANEDVCCSSYSWTGAUTPCAAERQKLPINALSNLSLRHNLNVTSTSRAC
QRKQVTFDRLOVLDSDHQDLREVKAASKVRKANLSVEEACSLTPHSKSKPFGYG
AKVDRHARKAQTHTINSWKDLLEDNVTPIIDTTIMAKNEVFCVQPEKGRKPARLIV
PDLGVRCERKALYDVVTKLPAYMGSSYFOYSPQORVEFLVQAKSKKTPMGFYD
TRCFDSTVIESDIETEAICYOCDLDPQARVAIKSLTERLYGVGGLTNSRGENGYRR
CRASGLVITSCGNTLTCYIKARAACRAAGLQDCTMLVCGDDLVLCIESAGVQDDAASL
RAFTAMTRYSAAPPDPOPEYDLELITSCSSNVVAHDCAGKRVYLLTPDPTPLAR
AAWETASHTPVNSHLGNILMFAPTLWARMILMTHTFSVLIAHQQLDCEIYGACY
SIEPLDLPPIIQRL"

BASE COUNT 1495 a 2218 c 2058 g 1539 t
ORIGIN

Alignment Scores:

Pred. No.: 4,5e-67 Length: 7310
Score: 943.00 Matches: 179
Percent Similarity: 100.00% Conservativity: 3
Best Local Similarity: 98.35% Mismatches: 0
Query Match: 98.95% Indels: 0
DB: 6 Gaps: 0

US-09-965-594-1 (1-182) x HPCPOLYP (1-7310)

QY 1 MetaAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20

```

:::|||||
1728 CTGGCGCCATACGGCGTACGCCACACAAAGGGCCCTCTAGGGTCATAATCAACC 1787
QY 21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyValGlnIleValSerThrAla 40
DB AGCCTAACTGCGCGGCAAAACCAAGTGAGGGTGAGGTCCAGATTGTGTCAACTGCT 2853
1788 AGCCTAACTGCGCGGCAAAACCAAGTGAGGGTGAGGTCCAGATTGTGTCAACTGCT 1847
QY 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
DB GCCCAAACTTCTCTGGCAACGTGCATCAATGGGGTGCTGGACTGTCTACCAACGGGGCC 2913
61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
DB GGAACGAGGACCATCGCGTCCACCAAGGGTCTCTCATCCAGATGTATACCAATGTAGAC 2973
1908 GCAACGAGGACCATCGCGTCCACCAAGGGTCTCTCATCCAGATGTATACCAATGTAGAC 1967
QY 81 LysAspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
DB CAAGACCTTGTGGGTGCGCGCTCCGCAAGGTAGCCGCTCAITGACACCTGCACCTGC 2027
101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 120
DB GGCTCTCGGACCTTTACCTGGTCACAGGACGCCGATGTCAITCCGTCGCGCGCGG 2087
121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
DB GGTGATACAGGGGACGCTGCTGCGCGCGCGCCATTTCCTACTTSAAGGCTGCTCC 2147
141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyTyrPheArgAlaAlaValCys 160
DB GGGGTCTCGCTGTGTGCGCGCGCGGACCGCTGGGCATATTAGGGCGCGGTGTC 2207
161 ThrArgGlyValAlaAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
DB ACCCGTGGAGTGGCTAAGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACCATG 2267
QY 181 ArgSer 182
DB |||||
2268 AGGTCC 2273

RESULT 14
AR118703
LOCUS AR118703 8316 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 88 from patent US 6150087.
ACCESSION AR118703
VERSION AR118703.1 GI:14100613
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 8316)
AUTHORS Chien,D.Y.
TITLE NANBV diagnostics and vaccines
JOURNAL Patent: US 6150087-A 88 21-NOV-2000;
FEATURES
    source
        1..8316
        /organism="unknown"
BASE COUNT 1671 a 2529 c 2345 g 1771 t
ORIGIN

Alignment Scores:
Pred. No.: 5,17e-67 Length: 8316
Score: 943.00 Matches: 179
Percent Similarity: 100.00% Conservative: 3
Best Local Similarity: 98.35% Mismatches: 0
Query Match: 98.95% Indels: 0
DB: 6 Gaps: 0

US-09-965-594-1 (1-182) x AR118703 (1-8316)

QY 1 MetaLapProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
DB ::|||
2734 CTGGCGCCCATCAGCGCGTACGCCAGCAGACAAAGGGCCCTCTAGGGTGCTATAATCACC 2793

```

```

QY 21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyValGlnIleValSerThrAla 40
DB |||||
2794 AGCCTAACTGCGCGGCAAAACCAAGTGAGGGTGAGGTCCAGATTGTGTCAACTGCT 2853
QY 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
DB |||||
2854 GCCCAAACTTCTCTGGCAACGTGCATCAATGGGGTGCTGGACTGTCTACCAACGGGGCC 2913
QY 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
DB |||||
2914 GGAACGAGGACCATCGCGTCCACCAAGGGTCTCTCATCCAGATGTATACCAATGTAGAC 2973
QY 81 LysAspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
DB ::|||
2974 CAAAGACCTTGTGGGTGCGCGCTCCGCAAGGTAGCCGCTCAITGACACCTGCACCTGC 3033
QY 101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 120
DB |||||
3034 GGTCTCTCGGACCTTTACCTGGTCACAGGACGCCGATGTCAITCCGTCGCGCGCGG 3093
QY 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
DB |||||
3094 GGTGATACAGGGGACGCTGCTGCGCGCGCGCCATTTCCTACTTGAAGGCTCCTCG 3153
QY 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyTyrPheArgAlaAlaValCys 160
DB |||||
3154 GGGGTCTCGCTGTGTGCGCGCGGACCGCTGGGCATATTAGGGCGCGGTGTC 3213
QY 161 ThrArgGlyValAlaAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
DB |||||
3214 ACCCGTGGAGTGGCTAAGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACCATG 3273
QY 181 ArgSer 182
DB |||||
3274 AGGTCC 3279

RESULT 15
AR118728
LOCUS AR118728 8987 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 137 from patent US 6150087.
ACCESSION AR118728
VERSION AR118728.1 GI:14100638
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 8987)
AUTHORS Chien,D.Y.
TITLE NANBV diagnostics and vaccines
JOURNAL Patent: US 6150087-A 137 21-NOV-2000;
FEATURES
    source
        1..8987
        /organism="unknown"
BASE COUNT 1807 a 2735 c 2547 g 1898 t
ORIGIN

Alignment Scores:
Pred. No.: 5,62e-67 Length: 8987
Score: 943.00 Matches: 179
Percent Similarity: 100.00% Conservative: 3
Best Local Similarity: 98.35% Mismatches: 0
Query Match: 98.95% Indels: 0
DB: 6 Gaps: 0

US-09-965-594-1 (1-182) x AR118728 (1-8987)

QY 1 MetaLapProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
DB ::|||
3076 CTGGCGCCCATCAGCGCGTACGCCAGCAGACAAAGGGCCCTCTAGGGTGCTATAATCACC 3135
QY 21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyValGlnIleValSerThrAla 40
DB |||||
3136 AGCCTAACTGCGCGGCAAAACCAAGTGAGGGTGAGGTCCAGATTGTGTCAACTGCT 3195

```



```
Qy      41  AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
Db      3196  GCCCAAAACCTTCCTGGCAAGTGCATCAATGGGGTGTGTGGACTGTCTACCACGGGGCC 3255
Qy      61  GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
Db      3256  GGAACGAGGACCATCGGTCACCCAAAGGGTCTGTGTCATCCAGATGTATACCAATGTAGAC 3315
Qy      81  LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerIleuThrProCysThrCys 100
Db      3316  CAAGACCTTGTGGGCTGGCCGCTCGCAAGGTAGCGCTCATTTGACACCTGCACCTTGC 3375
Qy     101  GlySerSerAspLeuTyrIleuValThrArgHisAlaAspValIleProValArgArg 120
Db      3376  GGCTCCTCGGACCTTTACCTGGTCAGAGCGCAGCGCGATGTCATCCCGTGGCGCGCGG 3435
Qy     121  GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
Db      3436  GGTGATAGCAGGGGCGAGCCTGCTGTGCGCCGCGCCCATTTCTACTTGAAGGCTCCTCG 3495
Qy     141  GlyGlyProLeuLeuCysProAlaGlyHisAlaValAlcIlyIlePheArgAlaAlaValCys 160
Db      3496  GGGGTCCTGCTGTGTGCCCCGGGGCGACCGCGTGGGCATATTAGGGCGCGGTGTGC 3555
Qy     161  ThrArgGlyValAlaLysAlaValAspPheIleProValIleuSerLeuGluThrThrMet 180
Db      3556  ACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACACCATG 3615
Qy     181  ArgSer 182
Db      3616  AGGTCC 3621
```

Search completed: August 31, 2003, 00:45:16
Job time : 2382.6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 17:42:58 ; Search time 41.2251 Seconds
(Without alignments)
700.745 Million cell updates/sec

Title: US-09-965-594-1

Perfect score: 953

Sequence: 1 MAPITAYAQTRGLGCIIT.....GVAKAVDFIPVESLETTMRS 182

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

```

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	953	100.0	182	21	AA15211 Hepatitis C virus
2	946	99.3	686	23	AA18689 HCV-1 NS3/4a mutant
3	946	99.3	686	23	AAU76377 Hepatitis C virus
4	946	99.3	686	24	ABG72261 HCV-1 NS3/4a confo
5	944	99.1	3011	14	AAK40120 HCV genomic amino
6	943	99.0	609	15	AAK51170 Hepatitis C virus
7	943	99.0	1766	10	AAK92041 Sequence encoded i
8	943	99.0	1786	10	AAK90158 Protein sequence o
9	943	99.0	2261	10	AAK90164 Peptide encoded by

10	943	99.0	2301	10	AAK92047	Sequence encoded i
11	943	99.0	2436	10	AAK92050	Sequence encoded i
12	943	99.0	2436	10	AAK90288	Peptide encoded by
13	943	99.0	2772	21	AAK18540	Protein encoded by
14	943	99.0	2816	14	AAK34009	HCV-1 polyprotein.
15	943	99.0	2894	16	AAK70230	Composite hepatiti
16	943	99.0	2955	20	AAK14975	Amino acid sequenc
17	943	99.0	2955	21	AAK18541	Polyprotein encode
18	943	99.0	3011	13	AAK21519	Compiled HCV seque
19	943	99.0	3011	14	AAK31621	Hepatitis C virus
20	943	99.0	3011	17	AAK90931	Hepatitis C virus
21	943	99.0	3011	18	AAK34480	HCV polyprotein.
22	943	99.0	3011	19	AAK40038	HCV polyprotein.
23	943	99.0	3011	23	AAK22049	Hepatitis C virus
24	943	99.0	3011	23	AAU84597	HCV polyprotein la
25	942	98.8	632	23	AAE21847	Hepatitis C virus
26	942	98.8	632	23	AAE19905	Hepatitis C virus
27	942	98.8	686	23	AAE21837	Hepatitis C virus
28	942	98.8	686	23	AAE21838	Hepatitis C virus
29	942	98.8	686	23	AAE21839	Hepatitis C virus
30	942	98.8	686	23	AAE21840	Hepatitis C virus
31	942	98.8	686	23	AAE21841	Hepatitis C virus
32	942	98.8	686	23	AAE21842	Hepatitis C virus
33	942	98.8	686	23	AAE21843	Hepatitis C virus
34	942	98.8	686	23	AAE21844	Hepatitis C virus
35	942	98.8	686	23	AAE21845	Hepatitis C virus
36	942	98.8	686	23	AAE21846	Hepatitis C virus
37	942	98.8	686	23	AAE19900	Hepatitis C virus
38	942	98.8	686	23	AAE19907	Hepatitis C virus
39	942	98.8	686	23	AAE19908	Hepatitis C virus
40	942	98.8	686	23	AAE19919	Hepatitis C virus
41	942	98.8	686	23	AAE19920	Hepatitis C virus
42	942	98.8	686	23	AAE19921	Hepatitis C virus
43	942	98.8	686	23	AAE19922	Hepatitis C virus
44	942	98.8	686	23	AAE19923	Hepatitis C virus
45	942	98.8	686	23	AAE19924	Hepatitis C virus

ALIGNMENTS

RESULT 1
AAB15211
ID AAB15211 standard; protein; 182 AA.
AC AAB15211;
XX
XX
XX 19-DEC-2000 (first entry)
DT
DE Hepatitis C virus NS3 protease.
XX
XX Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW liver failure; liver cancer.
XX
XX Hepatitis C virus.
XX
XX WO200040707-A1.
XX
XX 13-JUL-2000.
XX
XX 06-JAN-2000; 2000WO-US00345.
XX
XX 08-JAN-1999; 990S-0115271.
XX
XX (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX
XX WPI; 2000-465976/40.
XX
XX N-PSDB; AAA70344.
XX
XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
XX substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic

PT amino acid, useful for screening inhibitors that may treat hepatitis C
PT -
XX
PS
PS Claim 3; Fig 9; 66pp; English.
XX
XX The present sequence is the Hepatitis C virus (HCV) NS3 protease enzyme.
CC This protein is essential for the replication of the virus, acting to
CC cleave its replicative proteins from the polypeptide produced from the
CC HCV genome. NS4A is also needed for this process and inhibitors of the
CC two proteins should act as antiviral treatments of HCV infection. This is
CC useful as HCV can lead to chronic liver disease such as cirrhosis, liver
CC failure and liver cancer. The present invention concerns a number of NS3
CC mutants and NS3-NS4A fusion proteins which can be used to identify
CC inhibitors of this type as well as enabling structural studies of the
CC protease and protease-inhibitor complexes.
XX
SQ Sequence 182 AA;
Query Match 100.0%; Score 953; DB 21; Length 182;
Best Local Similarity 100.0%; Pred. No. 3.9e-91;
Matches 182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAPITAYAOQTGRLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTYYHGA 60
Db 1 MAPITAYAOQTGRLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTYYHGA 60
QY 61 GTRTIASPKGPVIOYNTVDKDLVGPAPQGSRSITPCTCGSSDLYLVTRHADVIPVRRR 120
Db 61 GTRTIASPKGPVIOYNTVDKDLVGPAPQGSRSITPCTCGSSDLYLVTRHADVIPVRRR 120
QY 121 GDSRGSLLSPRPISYILKSGSGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 180
Db 121 GDSRGSLLSPRPISYILKSGSGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 180
QY 181 RS 182
Db 181 RS 182
RESULT 2
AAE18689
ID AAE18689 standard; Protein: 686 AA.
XX
AC AAE18689;
XX
XX 17-MAY-2002 (first entry)
XX
XX HCV-1 NS3/4a mutant conformational antigen.
XX
XX Hepatitis C virus; NS3/4a antigen; HCV infection; mutant; mutein.
XX
XX Hepatitis C virus type 1.
OS Synthetic.
XX
XX Key Location/Qualifiers
FT Misc-difference 403 /note- "Wild type Thr substituted with Pro"
FT Misc-difference 404 /note- "Wild type Ser substituted with Ile"
FT
XX WO200196875-A2.
XX
XX 20-DEC-2001.
XX
XX 14-JUN-2001; 2001WO-US19369.
XX
XX 15-JUN-2000; 2000US-212082P.
PR 02-APR-2001; 2001US-280811P.
PR 02-APR-2001; 2001US-280867P.
XX
XX (CHIR) CHIRON CORP.
XX
XX Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Coit D;
PI

PI Medina-Selby A;
XX
DR WP1: 2002-179522/23.
DR N-PSDB; AAD29795.
XX
XX Immunoassay solid support useful for detecting hepatitis C virus
PT infection in a biological sample, comprises at least one of HCV
PT anti-core antibody and HCV NS3/4a epitope, bound to the support -
XX
PS Example 2; Fig 4; 87pp; English.
XX
XX The present invention relates to hepatitis C virus (HCV) core antigen
CC and NS (nonstructural) 3/4a antibody combination assay that can detect
CC both HCV antigens and antibodies present in a sample using a single
CC solid matrix as well as immunoassay solid supports for use in the assay.
CC The solid support is useful for detecting HCV infection in a biological
CC sample. The present sequence is HCV-1 NS3/4a mutant conformational
CC antigen. This sequence is used in the exemplification of the invention.
XX
SQ Sequence 686 AA;
Query Match 99.3%; Score 946; DB 23; Length 686;
Best Local Similarity 98.9%; Pred. No. 1.2e-89;
Matches 180; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAPITAYAOQTGRLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTYYHGA 60
Db 1 MAPITAYAOQTGRLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTYYHGA 60
QY 61 GTRTIASPKGPVIOYNTVDKDLVGPAPQGSRSITPCTCGSSDLYLVTRHADVIPVRRR 120
Db 61 GTRTIASPKGPVIOYNTVDKDLVGPAPQGSRSITPCTCGSSDLYLVTRHADVIPVRRR 120
QY 121 GDSRGSLLSPRPISYILKSGSGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 180
Db 121 GDSRGSLLSPRPISYILKSGSGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 180
QY 181 RS 182
Db 181 RS 182
RESULT 3
AAU76377
ID AAU76377 standard; Protein: 686 AA.
XX
AC AAU76377;
XX
XX 08-MAY-2002 (first entry)
XX
XX Hepatitis C virus NS3/4a conformational epitope protein sequence.
XX
XX Hepatitis C virus; HCV; NS3/4a conformational epitope; seroconversion;
XX immunoassay solid support; multiple epitope fusion antigen; MEFA;
XX non-structural protein; mutant; mutein.
XX
XX Hepatitis C virus.
OS Synthetic.
XX
XX Key Location/Qualifiers
FT Misc-difference 403 /note- "Wild-type Thr substituted by Pro"
FT Misc-difference 404 /note- "Wild-type Ser substituted by Ile"
FT
XX WO200196870-A2.
XX
XX 20-DEC-2001.
XX
XX 14-JUN-2001; 2001WO-US19156.
XX
XX 15-JUN-2000; 2000US-212082P.
PR 02-APR-2001; 2001US-280811P.
PR

```

PR 02-APR-2001; 2001US-280867P.
XX (CHIR ) CHIRON CORP.
PA Chien DY, Arcangel P, Tandeske L, George-nascimento C, Coit D;
PI Medina-selby A;
XX WPI: 2002-090228/12.
DR N-PSDB; ABK15344.
XX Immunoassay solid support, useful for detecting hepatitis C virus
PT infection in biological sample, comprises HCV NS3/4a conformational
PT epitope and multiple epitope fusion antigen bound to the support .
XX Claim 5; Fig 3; 92pp: English.
XX The present invention relates to a new immunoassay solid support
CC consisting essentially of at least one hepatitis C virus (HCV) NS3/4a
CC conformational epitope and a multiple epitope fusion antigen (MEFA),
CC bound to the support. The NS3/4a conformational epitope and/or
CC MEFA reacts specifically with anti-HCV antibodies present in a biological
CC sample from an HCV-infected individual. The immunoassay of the invention
CC is useful for detecting hepatitis C virus infection in a biological
CC sample. The method of the invention provides a sensitive, accurate
CC diagnostic and prognostic tool to provide adequate patient care and to
CC prevent transmission of HCV by blood and by blood products, or by
CC personal contact. Use of NS3/4a conformational epitope in combination
CC with MEFA, provides a sensitive and reliable method for detecting early
CC HCV seroconversion. Use of MEFA has the added advantages of decreasing
CC masking problems, improving sensitivity in detecting antibodies by
CC allowing a greater number of epitopes on a unit surface area of
CC substrate, and improving substrate. Detection accuracy is increased and
CC the incidence of false results is reduced because of the identification
CC and the use of highly immunogenic HCV antigens which are present during
CC the early stages of HCV seroconversion. The present amino acid sequence
CC represents the non-structural protein NS3/4a conformational epitope of
CC the invention.
XX Sequence 586 AA:
SQ
    Query Match          99.3%; Score 946; DB 23; Length 686;
    Best Local Similarity 98.9%; Pred. No. 1.2e-89;
    Matches 180; Conservative 2; Mismatches 0; Indels 0; Gaps 0:
QY 1 MAPITAYAOOTRGLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
DB 1 MAPITAYAOOTRGLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
QY 61 GTRTIASPKGPVIQMYTNVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 120
DB 61 GTRTIASPKGPVIQMYTNVDQDLVGMWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 120
QY 121 GDSRGLSLSPRIYSYKSGSGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVSLFTTM 180
DB 121 GDSRGLSLSPRIYSYKSGSGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVLENFTTM 180
QY 181 RS 182
DB 181 RS 182

```

RESULT 4

ABG72261
ID ABG72261 standard; Protein; 686 AA.

```

XX AC ABG72261;
XX AC
XX DT 06-MAR-2003 (first entry)
XX DE HCV-1 NS3/4a conformational antigen.
XX KW Immunoassay solid support; Hepatitis C Virus type-1; HCV-1;
XX NS3/4a conformational epitope; multiple epitope fusion antigen;
KW

```

```

KW MEFA: anti-HCV antibody; NS3/4a conformational antigen;
KW HCV infection; mutant; mutein.
XX Hepatitis C virus type 1.
OS Synthetic.
XX Key Location/Qualifiers
FT Region 2..686 "Corresponds to amino acid residues 1027-1711
FT of HCV-1 NS3/4a polypeptide"
FT Misc-difference 403 /note= "Substitution of wild-type Thr to Pro"
FT Misc-difference 404 /note= "Substitution of wild-type Ser to Ile"
XX US2002146685-A1.
XX 10-OCT-2002.
XX 14-JUN-2001; 2001US-0881654.
XX 15-JUN-2000; 2000US-212082P.
PR 02-APR-2001; 2001US-280811P.
PR 02-APR-2001; 2001US-280867P.
XX (CHIE/) CHIEN D Y.
PA (ARCA/) ARCANGEL P.
PA (TAND/) TANDESKE L.
PA (GEOR/) GEORGE-NASCIMENTO C.
PA (COIT/) COIT D.
PA (MEDI/) MEDINA-SELBY A.
XX Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Coit D;
PI Medina-Selby A;
XX WPI: 2003-147573/14.
DR N-PSDB; ABX14410.
XX

```

Immunoassay solid support for detecting Hepatitis C Virus infection in biological samples, comprises Hepatitis C Virus conformational epitope and multiple epitope fusion antigen -

Claim 2; Fig 3A-3D; 45pp: English.

The present invention relates to immunoassays comprising Hepatitis C virus (HCV) NS3/4a conformational epitope and multiple epitope fusion antigen (MEFA), bound to a solid support. The NS3/4a epitope and/or the multiple epitope fusion antigen react with anti-HCV antibodies present in a biological sample from an HCV-infected individual. The immunoassays and methods of the invention are useful for detecting HCV infection in a biological sample. The inventive immunoassay solid support provides a sensitive and reliable method for detecting early HCV seroconversion. The assays can detect HCV infection caused by any six known genotypes of HCV. The use of the multiple epitope fusion proteins decreases masking problems, improves sensitivity in detecting antibodies by allowing a greater number of epitopes on a unit area of substrate, and improves selectivity. The present sequence represents HCV type 1 (HCV-1) NS3/4a conformational antigen, a mutant of the HCV-1 NS3/4a polypeptide.

Sequence 686 AA;

Query Match 99.3%; Score 946; DB 24; Length 686;

Best Local Similarity 98.9%; Pred. No. 1.2e-89;

Matches 180; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPITAYAOOTRGLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60

DB 1 MAPITAYAOOTRGLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60

QY 61 GTRTIASPKGPVIQMYTNVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 120

DB 61 GTRTIASPKGPVIQMYTNVDQDLVGMWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 120

```

Qy 121 GDSRGLSPRPISYLKSGSGPLLCAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 180
Db 121 GDSRGLSPRPISYLKSGSGPLLCAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 180
Qy 181 RS 182
Db 181 RS 182

RESULT 5
AAR40120
ID AAR40120 standard; Protein: 3011 AA.
XX
AC AAR40120;
XX
DT 25-MAR-2003 (updated)
DT 27-JAN-1994 (first entry)
XX
DE HCV genomic amino acid sequence isolated from infected human LG.
XX
KW Hepatitis C Virus; Non-A, non-B hepatitis Virus; HCV; NANBHV;
KW human growth hormone; HGH; secretion signal; fusion protein;
KW vaccine.
XX
OS Hepatitis C Virus.
XX
PN WC9315193-A1.
XX
PD 05-AUG-1993.
XX
PF 29-JAN-1993; 93MO-US00907.
XX
PR 31-JAN-1992; 92US-0830024.
XX
PA (ABBO ) ABBOTT LAB.
XX
PI Bode SL, Casey JM, Desai SM, Devare SG, Frail DE;
PI Yamaguchi J, Zeck BJ;
XX
DR WPI; 1993-258673/32.
XX
PT New plasmid pHCV-162 is a mammalian expression systems for HCV
PT proteins - useful for diagnosing HCV infection and as vaccines
PT for preventing HCV infection
XX
PS Example 1; Page 39-49; 100pp: English.
XX
CC RNA was isolated from the plasma of a HCV seropositive human
CC (designated "LG") and cDNA was prepared from it. The cDNA was
CC PCR amplified using specific primers with sequences based
CC on the prototype HCV-1 cDNA sequence (GHBANK M82321). Further
CC amplification using nested primers resulted in 7 adjacent HCV DNA
CC fragments which could be assembled into a full-length sequence. The
CC DNA sequence was determined and translated into the genomic amino
CC acid sequence. Comparison of the LG genomic amino acid sequence
CC with that from HCV-1 showed 134 amino acid differences.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 3011 AA.

Query Match 99.1%; Score 944; DB 14; Length 3011;
Best Local Similarity 98.4%; Pred. No. 1.4e-88;
Matches 179; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAPITAYAOOTRGLGCIITSITGRDKNOVEGEVQIVSTAAQTFLATCINGVCWTYHGA 60
Db 1026 LAPITAYAOOTRGLGCIITSITGRDKNOVEGEVQIVSTAAQTFLATCINGVCWTYHGA 1085
Qy 61 GTRTIASPKGPVIOYNTVDKLVGHPAPQGSRLTPCTCGSSDLYLVTRHADVIPVRR 120
Db 1086 GTRTIASPKGPVIOYNTVDKLVGHPAPQGSRLTPCTCGSSDLYLVTRHADVIPVRR 1145

```

```

Qy 121 GDSRGLSPRPISYLKSGSGPLLCAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 180
Db 1146 GDSRGLSPRPISYLKSGSGPLLCAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 1205
Qy 181 RS 182
Db 1206 RS 1207

RESULT 6
AAR51170
ID AAR51170 standard; peptide: 609 AA.
XX
AC AAR51170;
XX
DT 20-OCT-1994 (first entry)
XX
DE Hepatitis C virus non-structural protein 3.
XX
KW Peptide; antibody; hepatitis C virus; HCV; identification;
KW diagnosis; non-A non-B hepatitis; NANB; detection.
XX
OS Hepatitis C virus.
XX
PN JP06056891-A.
XX
PD 01-MAR-1994.
XX
PF 05-AUG-1992; 92JP-0209201.
XX
PR 05-AUG-1992; 92JP-0209201.
XX
PA (OLYU ) OLYMPUS OPTICAL CO LTD.
XX
DR WPI; 1994-106803/13.
XX
PI New peptide(s) reactive with anti-hepatitis C virus antibody -
PI for specific, early diagnosis of HCV infection
XX
PS Disclosure; Page 9-10; 15pp; Japanese.
XX
CC Peptide fragments of the non-structural protein (NS3) are reactive
CC with and can detect antibodies against the NS3 domain of HCV. The
CC peptides can be used for diagnosis of HCV infection. Nonspecific
CC reaction can be inhibited and misdiagnosis of HCV infection can be
CC decreased. See AAR51162-70.
XX
SQ Sequence 609 AA.

Query Match 99.0%; Score 943; DB 15; Length 609;
Best Local Similarity 98.4%; Pred. No. 2.1e-89;
Matches 179; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAPITAYAOOTRGLGCIITSITGRDKNOVEGEVQIVSTAAQTFLATCINGVCWTYHGA 60
Db 20 LAPITAYAOOTRGLGCIITSITGRDKNOVEGEVQIVSTAAQTFLATCINGVCWTYHGA 79
Qy 61 GTRTIASPKGPVIOYNTVDKLVGHPAPQGSRLTPCTCGSSDLYLVTRHADVIPVRR 120
Db 80 GTRTIASPKGPVIOYNTVDKLVGHPAPQGSRLTPCTCGSSDLYLVTRHADVIPVRR 139
Qy 121 GDSRGLSPRPISYLKSGSGPLLCAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 180
Db 140 GDSRGLSPRPISYLKSGSGPLLCAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 199
Qy 181 RS 182
Db 200 RS 201

RESULT 7
AAP92041
ID AAP92041 standard; protein: 1765 AA.

```

```

XX AAP92041;
AC
XX
XX 25-MAR-2003 (updated)
DT 02-MAR-1990 (first entry)
XX
XX Sequence encoded in the hepatitis C virus (HCV) cDNA inserts in clones
DE 14i, 11b, 7f, 7e, 8h, 33c, 40b, 37b, 35, 36, 8i, 32, 33b, 25c, 14c, 8f,
DE 33f, 33g and 39c.
XX
XX Hepatitis C virus (HCV): non-A, non-B hepatitis (HANBH)
KW
XX Hepatitis C virus.
OS
XX
XX EP318216-A.
PN
XX
XX 31-MAY-1989.
PD
XX
XX 18-NOV-1988; 88EP-0310522.
PF
XX
XX 18-NOV-1987; 87US-0122714.
PR
XX 30-DEC-1987; 87US-0139886.
PR
XX 26-FEB-1988; 88US-0161072.
PR
XX 06-MAY-1988; 88US-0191263.
PR
XX 26-OCT-1988; 88US-0263584.
PR
XX 14-NOV-1988; 88US-0271450.
PR
XX (CHIR ) CHIRON CORP.
PA
XX
XX Houghton M, Choo QL, Kuo G;
PI
XX
XX WPI: 1989-159274/22.
DR
XX N-PSDB: AAN92097.
DR
XX
XX Purified hepatitis C virus
PT
XX - and associated nucleic acids and polypeptide(s)
PT
XX
XX Claim 13: Figure 26-1, 26-2, 26-3, 26-4, 26-5, 26-6; 139pp: English.
PS
XX
XX It is the sequence encoded in the open reading frame of hepatitis C virus
CC cDNA inserts in clones 14i, 11b, 7f, 7e, 8h, 33c, 40b, 37b, 35, 36,
CC 81, 32, 33b, 25c, 14c, 8f, a33f, 33g and 39c. It is antigenic and could
CC be used in immunoassay reagents and vaccines and to generate antibodies
CC useful in diagnosis and passive immunotherapy for HCV infection/non-A,
CC non-B hepatitis.
CC (Updated on 25-MAR-2003 to correct PR field.)
CC (Updated on 25-MAR-2003 to correct PI field.)
CC
XX
XX Sequence 1766 AA;
SQ
Query Match 99.0%; Score 943; DB 10; Length 1766;
Best Local Similarity 98.4%; Pred. No. 8.8e-89;
Matches 179; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MAPITAYAOOTRGLGCIITSLTGRDKNOVEGEVOIVSTAAOTFLATCINGVCWTVYHGA 60
Db 310 LAPITAYAOOTRGLGCIITSLTGRDKNOVEGEVOIVSTAAOTFLATCINGVCWTVYHGA 369
Qy 61 GRTTIASPKGPVIQMTNVYDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 120
Db 370 GRTTIASPKGPVIQMTNVYDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 429
Qy 121 GDSRGSLLSPRISYLYKSGSGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVRSLETTM 180
Db 430 GDSRGSLLSPRISYLYKSGSGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVRSLETTM 489
Qy 181 RS 182
Db 490 RS 491
RESULT 8
AAP90158
ID AAP90158 standard; protein; 1786 AA.
XX
AC AAP90158;
XX
XX 25-MAR-2003 (updated)
DT 10-NOV-1989 (first entry)
XX
XX Protein sequence of hepatitis c virus composite cDNA.
DE
XX
XX Hepatitis C virus; vaccine.
KW
XX Pan troglodytes.
OS
XX
XX GB2212511-A.
PN
XX
XX 26-JUL-1989.
PD
XX
XX 18-NOV-1988; 88GB-0027024.
PF
XX
XX 18-NOV-1987; 87US-0122714.
PR
XX 30-DEC-1987; 87US-0139886.
PR
XX 26-FEB-1988; 88US-0161072.
PR
XX 26-OCT-1988; 88US-0263584.
PR
XX (CHIR ) CHIRON CORPORATION.
PA
XX
XX Houghton M, Choo QL, Kuo G;
PI
XX
XX WPI: 1989-215054/30.
DR
XX N-PSDB: AAN90327.
DR
XX
XX Hepatitis C virus gene - used for prodn. of polynucleotide probes,
PT polypeptide(s) and antibodies for diagnosis, prevention and treatment
PT of infection.
XX
XX Disclosure; fig 26; 30pp: English.
PS
XX
XX The sequence is encoded by the composite cDNA of AAN90327. These
CC antigens react with antibodies in patients with non-A non-B hepatitis
CC (NANBH). They can be used to diagnose HCV-induced NANBH, to raise
CC antibodies for immunoassay or treatment, or to produce vaccines.
CC (Updated on 25-MAR-2003 to correct PR field.)
CC
XX
XX Sequence 1786 AA;
SQ
Query Match 99.0%; Score 943; DB 10; Length 1786;
Best Local Similarity 98.4%; Pred. No. 8.9e-89;
Matches 179; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MAPITAYAOOTRGLGCIITSLTGRDKNOVEGEVOIVSTAAOTFLATCINGVCWTVYHGA 60
Db 310 LAPITAYAOOTRGLGCIITSLTGRDKNOVEGEVOIVSTAAOTFLATCINGVCWTVYHGA 369
Qy 61 GRTTIASPKGPVIQMTNVYDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 120
Db 370 GRTTIASPKGPVIQMTNVYDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 429
Qy 121 GDSRGSLLSPRISYLYKSGSGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVRSLETTM 180
Db 430 GDSRGSLLSPRISYLYKSGSGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVRSLETTM 489
Qy 181 RS 182
Db 490 RS 491
RESULT 9
AAP90164
ID AAP90164 standard; protein; 2261 AA.
XX
XX AAP90164;
XX
XX 25-MAR-2003 (updated)
DT

```

DT 01-NOV-1989 (first entry)
 XX Peptide encoded by composite hepatitis C virus cDNA.
 DE Hepatitis C virus: clone 12f; clone 15e; probe; vaccine.
 XX Pan troglodytes.

OS GB2212511-A.
 XX 26-JUL-1989.
 PN 18-NOV-1988; 88GB-0027024.
 PD 18-NOV-1987; 87US-0122714.
 XX 30-DEC-1987; 87US-0139886.
 PF 26-FEB-1988; 88US-0161072.
 XX 26-OCT-1988; 88US-0263584.
 PR 26-FEB-1988; 88US-0161072.
 PR 26-OCT-1988; 88US-0263584.
 XX (CHIR) CHIRON CORPORATION.

XX Houghton M, Choo QL, Kuo G;
 XX WPI: 1989-215054/30.
 DR N-PSDB: AAN90331.

XX Hepatitis C virus gene - used for prodn. of polynucleotide probes,
 PT polypeptide(s) and antibodies for diagnosis, prevention and
 PT treatment of infection.
 XX Disclosure: fig 32; 235pp; English.

XX The sequence is the peptide encoded by the composite hepatitis C
 CC virus (HCV) cDNA of AAN90331. The polypeptides are used to diagnose
 CC HCV-induced NANBH, to raise antibodies for immunoassay or treatment,
 CC or to produce vaccines.
 CC (Updated on 25-MAR-2003 to correct PR field.)
 XX

SQ Sequence 2261 AA;
 Query Match 99.0%; Score 943; DB 10; Length 2261;
 Best Local Similarity 98.4%; Pred. No. 1.2e-88;
 Matches 179; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 MAPITAYAOOTRGLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTYHGA 60
 DB 401 LAPITAYAOOTRGLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTYHGA 460
 OY 61 GTRTIASPKGPVIOMYTNVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 120
 DB 461 GTRTIASPKGPVIOMYTNVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 520
 OY 121 GDSRGLSPRPISYKLGSSGGPILCPAGHANGIFRAAVCTRGVAKAVDFIPVESLETTM 180
 DB 521 GDSRGLSPRPISYKLGSSGGPILCPAGHANGIFRAAVCTRGVAKAVDFIPVESLETTM 580
 OY 181 RS 182
 DB 581 RS 582

RESULT 10
 AAP92047
 ID AAP92047 standard; protein: 2301 AA.
 XX
 AC AAP92047;

XX 25-MAR-2003 (updated)
 DT 02-MAR-1990 (first entry)
 XX
 DE Sequence encoded in the hepatitis C virus (HCV) cDNA inserts in clones
 DE 12f through 15e.
 XX

KW Hepatitis C virus (HCV); non-A, non-B hepatitis (HANBH).

XX Hepatitis C virus.

PN EP318216-A.

XX 31-MAY-1989.

PD 18-NOV-1988; 88EP-0310922.

XX 18-NOV-1987; 87US-0122714.

PR 30-DEC-1987; 87US-0139886.

PR 26-FEB-1988; 88US-0161072.

PR 06-MAY-1988; 88US-0191263.

PR 26-OCT-1988; 88US-0263584.

PR 14-NOV-1988; 88US-0271450.

XX (CHIR) CHIRON CORP.

XX Houghton M, Choo QL, Kuo G;
 XX WPI: 1989-159274/22.
 DR N-PSDB: AAN92103.

XX Purified hepatitis C virus
 PT - and associated nucleic acids and polypeptide(s)

XX Claim 13; Figure 32-1 - 32-7; 139 pp; English.

XX It is the sequence encoded in the open reading frame of hepatitis C virus
 CC (HCV) cDNA inserts in clones 12f through 15e. It is antigenic and could
 CC be used in immunoassay reagents and vaccines and to generate antibodies
 CC useful in diagnosis and passive immunotherapy for HCV infection/non-A,
 CC non-B hepatitis.
 CC (Updated on 25-MAR-2003 to correct PR field.)
 CC (Updated on 25-MAR-2003 to correct PI field.)
 XX

SQ Sequence 2301 AA;

Query Match 99.0%; Score 943; DB 10; Length 2301;
 Best Local Similarity 98.4%; Pred. No. 1.2e-88;
 Matches 179; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 MAPITAYAOOTRGLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTYHGA 60
 DB 401 LAPITAYAOOTRGLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTYHGA 460
 OY 61 GTRTIASPKGPVIOMYTNVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 120
 DB 461 GTRTIASPKGPVIOMYTNVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 520
 OY 121 GDSRGLSPRPISYKLGSSGGPILCPAGHANGIFRAAVCTRGVAKAVDFIPVESLETTM 180
 DB 521 GDSRGLSPRPISYKLGSSGGPILCPAGHANGIFRAAVCTRGVAKAVDFIPVESLETTM 580
 OY 181 RS 182
 DB 581 RS 582

RESULT 11
 AAP92050
 ID AAP92050 standard; protein: 2436 AA.
 XX
 AC AAP92050;

XX 25-MAR-2003 (updated)
 DT 02-MAR-1990 (first entry)
 XX
 DE Sequence encoded in the hepatitis C virus (HCV) cDNA inserts in clones
 DE K9-1 through 15e.
 XX Hepatitis C virus (HCV); non-A, non-B hepatitis (HANBH)

```

XX OS Hepatitis C virus.
XX PN EP18216-A.
XX XX
XX PD 31-MAY-1989.
XX PF 18-NOV-1988; 88EP-0310922.
XX PR 18-NOV-1987; 87US-0122714.
XX PR 30-DEC-1987; 87US-0139886.
XX PR 26-FEB-1988; 88US-0161072.
XX PR 06-MAY-1988; 88US-0191263.
XX PR 26-OCT-1988; 88US-0263584.
XX PR 14-NOV-1988; 88US-0271450.
XX PA (CHIR ) CHIRON CORP.
XX PI Houghton M, Choo QL, Kuo G;
XX DR N-PSDB; AAN92106.
XX DT WPI: 1989-159274/22.
XX DR N-PSDB; AAN92106.
XX XX
XX PT Purified hepatitis C virus
XX PT - and associated nucleic acids and polypeptide(s)
XX PS Claim 13; Figure 47-1 - 47-8; 139 pp; English.
XX CC It is the sequence encoded in the open reading frame of hepatitis C virus
XX CC (HCV) cDNA inserts in clones K9-1 through 15e. It is antigenic and could
XX CC be used in immunoassay reagents and vaccines and to generate antibodies
XX CC useful in diagnosis and passive immunotherapy for HCV infection/non-A,
XX CC non-B hepatitis.
XX CC (Updated on 25-MAR-2003 to correct PR field.)
XX CC (Updated on 25-MAR-2003 to correct PI field.)
XX SQ Sequence 2436 AA;

Query Match 99.0%; Score 943; DB 10; Length 2436;
Best Local Similarity 98.4%; Pred. No. 1.3e-88;
Matches 179; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPITAAQOTRGGLGCIITSLTGDKNOVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
DB 576 IAPITAAQOTRGGLGCIITSLTGDKNOVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 635
QY 61 GTRTIASPKGPVIQMYTNVDKDLVGPAPQGSRSLSLPTCTGSSDLVLTTRHADVIPVRRR 120
DB 636 GTRTIASPKGPVIQMYTNVDKDLVGPAPQGSRSLSLPTCTGSSDLVLTTRHADVIPVRRR 695
QY 121 GDSRGSLLSPRPISYLKSGSGPLLCPCAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 180
DB 696 GDSRGSLLSPRPISYLKSGSGPLLCPCAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 755
QY 181 RS 182
DB 756 RS 757

RESULT 12
AAP90288
ID AAP90288 standard; protein; 2436 AA.
XX AC AAP90288;
XX XX
XX XX 25-MAR-2003 (updated)
XX DT 19-JUL-2001 (updated)
XX DT 01-NOV-1989 (first entry)
XX XX
XX DE Peptide encoded by composite hepatitis C cDNA.
XX XX Hepatitis C virus; clone 15e; clone k9-1; probe; vaccine.
XX KW
XX XX

```

```

OS Pan troglodytes.
XX CB2212511-A.
XX PD 26-JUL-1989.
XX PF 18-NOV-1988; 88GB-0027024.
XX PR 18-NOV-1987; 87US-0122714.
XX PR 30-DEC-1987; 87US-0139886.
XX PR 26-FEB-1988; 88US-0161072.
XX PR 26-OCT-1988; 88US-0263584.
XX PA (CHIR ) CHIRON CORPORATION.
XX PI Houghton M, Choo QL, Kuo G;
XX DR N-PSDB; AAN90336.
XX DT WPI: 1989-215054/30.
XX DR N-PSDB; AAN90336.
XX XX
XX PT Hepatitis C virus gene - used for prodn. of polynucleotide probes,
XX PT polypeptide(s) and antibodies for diagnosis, prevention and
XX PT treatment of infection.
XX PS Disclosure: fig 47-1 to 47-8; 235pp; English.
XX CC The sequence is the peptide encoded by the composite hepatitis C
XX CC virus (HCV) cDNA of AAN90336. The polypeptides are used to
XX CC diagnose HCV-induced NANBH, to raise antibodies for
XX CC immunoassay or treatment, or to produce vaccines.
XX CC (N.B. This record was resubmitted to correct errors in the sequence.)
XX CC (Updated on 25-MAR-2003 to correct PR field.)
XX SQ Sequence 2436 AA;

Query Match 99.0%; Score 943; DB 10; Length 2436;
Best Local Similarity 98.4%; Pred. No. 1.3e-88;
Matches 179; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPITAAQOTRGGLGCIITSLTGDKNOVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
DB 576 LAPITAAQOTRGGLGCIITSLTGDKNOVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 635
QY 61 GTRTIASPKGPVIQMYTNVDKDLVGPAPQGSRSLSLPTCTGSSDLVLTTRHADVIPVRRR 120
DB 636 GTRTIASPKGPVIQMYTNVDKDLVGPAPQGSRSLSLPTCTGSSDLVLTTRHADVIPVRRR 695
QY 121 GDSRGSLLSPRPISYLKSGSGPLLCPCAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 180
DB 696 GDSRGSLLSPRPISYLKSGSGPLLCPCAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 755
QY 181 RS 182
DB 756 RS 757

RESULT 13
AAB18540
ID AAB18540 standard; Protein; 2772 AA.
XX AC AAB18540;
XX XX
XX XX 15-JAN-2001 (first entry)
XX DT
XX DE Protein encoded by a cDNA compiled Hepatitis C virus cDNA clones.
XX XX Hepatitis C virus; HCV; antisense polynucleotide; polypeptide;
XX KW viral infectivity; viral replication.
XX XX
XX OS Hepatitis C virus.
XX PN EP1034785-A2.
XX XX

```


PD 13-SEP-2000.
 XX
 PF 16-MAR-1990; 2000EP-0109602.
 XX
 XX 17-MAR-1989; 89US-0325338.
 PR 20-APR-1989; 89US-0341334.
 PR 18-MAY-1989; 89US-0350002.
 PR 16-MAR-1990; 90EP-0302866.
 XX
 PA (CHIR) CHIRON CORP.
 XX
 XX Houghton M, Choo Q, Kuo G;
 PI
 XX WPI; 2000-566891/53.
 DR N-PSDB; AAR75296.
 DR
 XX Novel composition comprising a hepatitis C virus antisense
 PT polynucleotide which is complementary to or corresponds to a sense
 PT strand of the virus genome, and selectively hybridizes to it.
 XX
 XX Example; Fig 16; 75pp; English.
 XX
 CC The specification describes a pharmaceutical composition which
 CC comprises a hepatitis C virus (HCV) antisense polynucleotide. The
 CC HCV is characterized by a positive stranded RNA genome which has
 CC 40% homology at the polypeptide level to a HCV polypeptide. The
 CC antisense polynucleotide binds to cellular polynucleotides which
 CC enhance and/or are required for viral infectivity, replicative
 CC ability or chronicity. The antisense polynucleotides may also be
 CC designed to bind with high specificity, to be of increased stability,
 CC to be stable and to have low toxicity. The composition also comprises
 CC an agent which causes viral RNA to be inactive. The composition
 CC is used for preventing HCV replication in a system. The present
 CC sequence is encoded by a novel HCV cDNA sequence, which is used in the
 CC course of the invention.
 XX
 SQ Sequence 2772 AA;

Query Match 99.0%; Score 943; DB 21; Length 2772;
 Best Local Similarity 98.4%; Pred. No. 1.6e-88;
 Matches 179; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MAPITAAQOTRGLGCIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTYHGA 60
 DB 912 LAPITAAQOTRGLGCIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTYHGA 971
 QY 61 GTRTIASPKGPVIOMYTNVDKDLVGMWPAPOGSRSLTPTCTGSSDLYLVTRHADVIPVRR 120
 DB 972 GTRTIASPKGPVIOMYTNVDKDLVGMWPAPOGSRSLTPTCTGSSDLYLVTRHADVIPVRR 1031
 QY 121 GDSRGSLLSPRPISYLGKSSGGPLLCPCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 180
 DB 1032 GDSRGSLLSPRPISYLGKSSGGPLLCPCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 1091
 QY 181 RS 182
 DB 1092 RS 1093

RESULT 14
 AAR34009
 ID AAR34009 standard; Protein; 2816 AA.
 XX
 AC AAR34009;
 XX
 XX 25-MAR-2003 (updated)
 DT 26-JUL-1993 (first entry)
 XX
 DE HCV-1 polypeptide.
 XX
 KW Polymerase chain reaction; PCR; amplifiy; primer; hepatitis C virus;
 KW HCV; asymptomatic; chronically infected; epitope; viral isolate;
 KW domain; immunological; cross-reactive; envelope protein; vaccine;

KW gp53(BVDV)/gp55; hog cholera virus; pestivirus; NSI; flavivirus.
 XX Hepatitis C virus.
 OS
 XX WO9306126-A1.
 XX
 PD 01-APR-1993.
 XX
 XX 11-SEP-1992; 92WO-US07683.
 PF
 XX 13-SEP-1991; 91US-0759575.
 PR
 XX (CHIR) CHIRON CORP.
 PA
 XX Houghton M, Weiner AJ;
 PI
 XX WPI; 1993-117468/14.
 DR
 XX Immuno-reactive hepatitis C virus polypeptide compsns. - contg.
 PT at least 2 sequences from the first variable domain of distinct
 PT HCV isolates
 XX
 PS Disclosure; Fig 9; 106pp; English.
 CC
 CC This sequence represents the entire hepatitis C virus polypeptide.
 CC HCV is a member of the flavivirus family and appears to encode a basic
 CC polypeptide domain ("C") at the N-terminal of the viral polypeptide.
 CC followed by two glycoprotein domains ("E1", "E2/NS1") upstream of the
 CC nonstructural genes NS2 through NS5. See also AAO39134-48, AAR33982-
 CC 4008 and AAR38088-89.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 CC
 SQ Sequence 2816 AA;

Query Match 99.0%; Score 943; DB 14; Length 2816;
 Best Local Similarity 98.4%; Pred. No. 1.6e-88;
 Matches 179; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MAPITAAQOTRGLGCIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTYHGA 60
 DB 1036 LAPITAAQOTRGLGCIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTYHGA 1085
 QY 61 GTRTIASPKGPVIOMYTNVDKDLVGMWPAPOGSRSLTPTCTGSSDLYLVTRHADVIPVRR 120
 DB 1086 GTRTIASPKGPVIOMYTNVDKDLVGMWPAPOGSRSLTPTCTGSSDLYLVTRHADVIPVRR 1145
 QY 121 GDSRGSLLSPRPISYLGKSSGGPLLCPCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 180
 DB 1146 GDSRGSLLSPRPISYLGKSSGGPLLCPCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 1205
 QY 181 RS 182
 DB 1206 RS 1207

RESULT 15
 AAR70230
 ID AAR70230 standard; protein; 2894 AA.
 XX
 AC AAR70230;
 XX
 XX 25-MAR-2003 (updated)
 DT 07-NOV-1995 (first entry)
 XX
 DE Composite hepatitis C virus (HC-J1/CDC/CHI).
 XX
 KW Composite hepatitis C virus; HC-J1/CDC/CHI; HCV; non-A non-B;
 KW synthetic antigens; blood screening.
 XX
 OS Hepatitis C virus.
 XX
 PN EP644202-A1.
 XX

PD 22-MAR-1995.
XX
XX 14-DEC-1990; 94EP-0108611.
XX
PR 14-DEC-1990; 90EP-0124241.
PR 14-DEC-1990; 90EP-0108611.
XX
XX (INNO-) INNOGENETICS NV.
XX
PI Deleys RJ, Maertens G, Pollet D, Van Heuverswyn H;
XX WPI: 1995-116946/16.
DR
XX Synthetic antigens for the detection of hepatitis C virus
PT antibodies - comprise portions of the Hcv peptide sequence, for
PT use in screening blood and blood products
XX
PS Disclosure; Fig 1; 5lpp; English.
XX
CC AAR70230 is the composite hepatitis C virus (HC-J1/CDC/CH1) protein
CC from which the synthetic HCV antigens described in AAR70210-R70229
CC were derived. These synthetic antigens can be used to screen blood,
CC or blood products for the presence HCV, they can also be used in
CC various specific assays for the detection of HCV antibodies, and
CC antigens, or as immunogens.
CC (Updated on 25-MAR-2003 to correct PN field.)
CC (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ Sequence 2894 AA;

Query Match 99.0%; Score 943; DB 16; Length 2894;
Best Local Similarity 98.4%; Pred. NO. 1.7e-88;
Matches 179; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPITAYACQTRGLGCIITSITGRKKNQVEGEVOIVSTAAQTFLATCINGVCTVYHGA 50
Db :|||||
1026 LAPITAYACQTRGLGCIITSITGRKKNQVEGEVOIVSTAAQTFLATCINGVCTVYHGA 1085

QY 61 GTRTIASPKGPVIOHYTNVNDKDLVGPAPQGSRSITPCTCGSSDLYLVTRHADVIPVRRR 120
Db |||||
1086 GTRTIASPKGPVIOHYTNVNDKDLVGPAPQGSRSITPCTCGSSDLYLVTRHADVIPVRRR 1145

QY 121 GDSRGSLLSPRISYILKSSGGPLLCPCAGHAGVGFRAAVCTRGVAKAVDFIPVESLETTM 180
Db |||||
1146 GDSRGSLLSPRISYILKSSGGPLLCPCAGHAGVGFRAAVCTRGVAKAVDFIPVENLETTM 1205

QY 181 RS 182
Db ||
1206 RS 1207

Search completed: August 30, 2003, 19:12:19
Job time : 42.2251 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 30, 2003, 19:13:57 ; Search time 169.009 Seconds
(without alignments)
2906.924 Million cell updates/sec

Title: US-09-965-594-1

Perfect score: 953

Sequence: 1 MAPITAYAQTRGLGCIIT.....GVAXAVDFIPVESLETIMRS 182

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=framer_p2n.model -p2n.model -DEV=xlip
-Q/cqn2.1/USPTO_spool/US09965594/runat_29082003_151918_28302/app_query_fasta_1.2872
-DB=N_Geneseq_19Jun03 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOPGC=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS-human40.cdi
-LIST=45 -DOCLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptco -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09965594.ecgn_1_1_1412 -runat_29082003_151918_28302 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSHLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -DELOP=6 -DEEXT=7

Database : N_Geneseq_19Jun03:*

1: /SIDSL/cgqdata/geneseq/geneseq-emb1/NA1980.DAT:*
2: /SIDSL/cgqdata/geneseq/geneseq-emb1/NA1981.DAT:*
3: /SIDSL/cgqdata/geneseq/geneseq-emb1/NA1982.DAT:*
4: /SIDSL/cgqdata/geneseq/geneseq-emb1/NA1983.DAT:*
5: /SIDSL/cgqdata/geneseq/geneseq-emb1/NA1984.DAT:*
6: /SIDSL/cgqdata/geneseq/geneseq-emb1/NA1985.DAT:*
7: /SIDSL/cgqdata/geneseq/geneseq-emb1/NA1986.DAT:*
8: /SIDSL/cgqdata/geneseq/geneseq-emb1/NA1987.DAT:*
9: /SIDSL/cgqdata/geneseq/geneseq-emb1/NA1988.DAT:*
10: /SIDSL/cgqdata/geneseq/geneseq-emb1/NA1989.DAT:*
11: /SIDSL/cgqdata/geneseq/geneseq-emb1/NA1990.DAT:*
12: /SIDSL/cgqdata/geneseq/geneseq-emb1/NA1991.DAT:*
13: /SIDSL/cgqdata/geneseq/geneseq-emb1/NA1992.DAT:*
14: /SIDSL/cgqdata/geneseq/geneseq-emb1/NA1993.DAT:*
15: /SIDSL/cgqdata/geneseq/geneseq-emb1/NA1994.DAT:*
16: /SIDSL/cgqdata/geneseq/geneseq-emb1/NA1995.DAT:*
17: /SIDSL/cgqdata/geneseq/geneseq-emb1/NA1996.DAT:*
18: /SIDSL/cgqdata/geneseq/geneseq-emb1/NA1997.DAT:*
19: /SIDSL/cgqdata/geneseq/geneseq-emb1/NA1998.DAT:*
20: /SIDSL/cgqdata/geneseq/geneseq-emb1/NA1999.DAT:*
21: /SIDSL/cgqdata/geneseq/geneseq-emb1/NA2000.DAT:*
22: /SIDSL/cgqdata/geneseq/geneseq-emb1/NA2001A.DAT:*
23: /SIDSL/cgqdata/geneseq/geneseq-emb1/NA2001B.DAT:*
24: /SIDSL/cgqdata/geneseq/geneseq-emb1/NA2002.DAT:*
25: /SIDSL/cgqdata/geneseq/geneseq-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	953	100.0	549	21	Hepatitis C virus
2	946	99.3	2058	24	HCV-1 NS3/4a mutan
3	946	99.3	2058	24	Hepatitis C virus
4	946	99.3	2058	25	DNA encoding HCV-1
5	943	99.0	5300	10	Combined open read
6	943	99.0	5360	10	Hepatitis C virus
7	943	99.0	6905	10	Combined open read
8	943	99.0	7310	10	Combined open read
9	943	99.0	7310	10	Composite hepatitis
10	943	99.0	7310	16	Hepatitis C virus
11	943	99.0	8316	21	cdna sequence comp
12	943	99.0	9133	20	Nucleotide sequenc
13	943	99.0	9185	11	Sense strand of th
14	943	99.0	9185	12	Hepatitis C virus
15	943	99.0	9185	21	Sense strand of HC
16	943	99.0	9400	13	Compiled HCV CDNA.
17	943	99.0	9401	17	Hepatitis C virus
18	943	99.0	9401	18	HCV polyprotein co
19	943	99.0	9401	19	HCV polyprotein co
20	943	99.0	9401	24	Hepatitis C virus
21	942	98.8	2061	24	Hepatitis C virus
22	942	98.8	2061	24	Hepatitis C virus
23	940	98.6	9185	20	Nucleotide sequenc
24	940	98.6	9185	20	Hepatitis C virus
25	939	98.5	8316	11	Hepatitis C virus
26	939	98.5	9502	15	Hepatitis C virus
27	937	98.3	9646	19	Hepatitis C virus
28	937	98.3	9646	24	cdna encoding hepa
29	937	98.3	12980	19	Hepatitis C virus
30	937	98.3	12980	24	Hepatitis C virus
31	937	98.3	16622	21	Nucleotide sequenc
32	936	98.2	630	17	Plasmid pT5His/HIV
33	936	98.2	630	17	HCV insoluble NS3
34	936	98.2	630	18	HCV NS3 protease C
35	936	98.2	810	17	PMB182delta4A HT e
36	936	98.2	810	17	HCV solubilised NS
37	936	98.2	810	18	HCV soluble NS3 pr
38	935	98.1	6299	22	HCV NS3A ORF comp
39	934	98.0	1933	20	HCV NS3 DNA. Hepa
40	934	98.0	8145	20	Plasmid pET-BS(+)
41	933	97.9	594	21	Hepatitis C virus
42	933	97.9	9365	24	Hepatitis C virus
43	933	97.9	9401	17	Hepatitis C virus
44	933	97.9	9416	19	Hepatitis C virus
45	933	97.9	9416	24	cdna encoding hepa

ALIGNMENTS

RESULT 1
AAA70344
ID AAA70344 standard; DNA; 549 bp.
XX
XX AAA70344;
AC
XX
DT 19-DEC-2000 (first entry)
XX
XX Hepatitis C virus NS3 protease coding sequence.
DE
DE Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW Liver failure; liver cancer; ds.
XX
XX Hepatitis C virus.
OS
XX
XX Key Location/Qualifiers
FT CDS 1..549

```

FI      /*tag- a
FT      /product= "NS3 protease"
PN      WO200040707-A1.
PX      13-JUL-2000.
PD      06-JAN-2000; 2000WO-US00345.
PX      08-JAN-1999; 99US-0115271.
XX      (BRIM ) BRISTOL-MYERS SQUIBB CO.
PA      Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
PI      WPI: 2000-465976/40.
XX      P-PSDB; AABIS211.
DR      Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
PT      substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
PT      amino acid, useful for screening inhibitors that may treat hepatitis C
PT      .
XX      Example 1: Fig 9; 65pp; English.
PS      The present sequence is the coding sequence for the Hepatitis C virus
XX      (HCV) NS3 protease enzyme. This protein is essential for the replication
CC      of the virus, acting to cleave its replicative proteins from the
CC      polyprotein produced from the HCV genome. NS4A is also needed for this
CC      process and inhibitors of the two proteins should act as antiviral
CC      treatments of HCV infection. This is useful as HCV can lead to chronic
CC      liver disease such as cirrhosis, liver failure and liver cancer. The
CC      present invention concerns a number of NS3 mutants and NS3-NS4A fusion
CC      proteins which can be used to identify inhibitors of this type, as well
CC      as enabling structural studies of the protease and protease:inhibitor
CC      complexes.
XX      Sequence 549 BP; 83 A; 180 C; 142 G; 144 T; 0 other;

Alignment Scores:
Pred. No.:      6.35e-80      Length:      549
Score:          953.00      Matches:      182
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    21      Indels:      0
DB:             21      Gaps:      0

US-09-965-594-1 (1-182) x AAA70344 (1-549)
QY      1 MetAlaProIleThrAlaTyAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
DB      1 ATGGCTCGGATCACCGCTTACGCTCAGCAGACCGTGGTCTGGGTGCATCATCACC 60
QY      21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
DB      61 TCCCTGACCGGTGCTGACAAACACAGGTTCAGGTGAGTTCAGATCGTTTCCACCGCT 120
QY      41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysIleThrValThrHisGlyAla 60
DB      121 GCTCAGACCTTCTCGGTACCTGCAACACGGTGTTCCTGGACCGTTTACACAGGTGCT 180
QY      61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAspValAsp 80
DB      181 GGTACCGGTACCATCGCTTCCCGAAAGGTCCGGTTATCCAGATGTACACCAAGTTGAC 240
QY      81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
DB      241 AAAGACCTGGTTGGTGGCGCTCCGACAGGTTCCTCGTTCCCTGACCCGCTGCACCTGC 300
QY      101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 120
DB      301 GGTTCCTCGACCTGTACTGTGTTACCGTACGCTGACGTATATCCCGGTTCGTCGCTGT 360

```

```

QY      121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
DB      361 GGTGACTCCCGTGGTTCCTGCTCCCGCGCTCCGATCTCCTACCTGAAAGGTTCCTCC 420
QY      141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
DB      421 GGTGTCGCTGCTGTGCGCGGTGGTCACGCTGTGGTATCTTCGCTGCTGCTGTTGC 480
QY      161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
DB      481 ACCCGTGGTGTGCTAAAGCTGTTGACTTCATCCGCTTGAATCCCTGGAACACCATG 540
QY      181 ArgSer 182
DB      541 CGTTCC 546
RESULT 2
AAD29795
ID      AAD29795 standard: DNA; 2058 BP.
XX      AC      AAD29795;
XX      DT      17-MAY-2002 (first entry)
XX      DE      HCV-1 NS3/4a mutant conformational antigen encoding DNA.
XX      KW      Hepatitis C virus; NS3/4a antigen; HCV infection; mutant; ds.
OS      Hepatitis C virus type 1.
XX      FH      Key      Location/Qualifiers
FT      CDS      1..686      /*tag= a
FT      /product= "HCV-1 NS3/4a conformational antigen"
FT      /note= "CDS does not include stop codon"
FT      /partial
XX      W0200196875-A2.
XX      PD      20-DEC-2001.
XX      PF      14-JUN-2001; 2001WO-US19369.
XX      PR      15-JUN-2000; 2000US-212082P.
XX      PR      02-APR-2001; 2001US-280811P.
XX      PR      02-APR-2001; 2001US-280867P.
XX      (CHIR ) CHIRON CORP.
XX      Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Coit D;
PI      Medina-Selby A;
XX      WPI: 2002-179522/23.
DR      P-PSDB; AAE18689.
XX      Immunoassay solid support useful for detecting hepatitis C virus
PT      infection in a biological sample, comprises at least one of HCV
PT      anti-core antibody and HCV NS3/4a epitope, bound to the support
XX      Example 2; Fig 4; 87pp; English.
XX      The present invention relates to hepatitis C virus (HCV) core antigen
CC      and NS (nonstructural) 3/4a antibody combination assay that can detect
CC      both HCV antigens and antibodies present in a sample using a single
CC      solid matrix as well as immunoassay solid supports for use in the assay.
CC      The solid support is useful for detecting HCV infection in a biological
CC      sample. The present sequence is a DNA encoding HCV-1 NS3/4a mutant
CC      conformational antigen. This sequence is used in the exemplification
CC      of the invention.
XX      Sequence 2058 BP; 419 A; 634 C; 580 G; 425 T; 0 other;
SQ

```

Alignment Scores:

```

Pred. No.:      1 46e-78      Length:      2058
Score:          946.00      Matches:      180
Percent Similarity: 100.00%      Conservative: 2
Best Local Similarity: 98.90%      Mismatches: 0
Query Match:      99.27%      Indels: 0
DB:              24      Gaps: 0

US-09-965-594-1 (1-182) x AAD29795 (1-2058)

QY 1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
DB 1 ATGGCGCCCATCACGGCGTAGCCGAGCAGACAAAGGGCCCTCTAGGTGCAATAATCAC 60
QY 21 SerLeuThrGlyArgAspGlyAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
DB 61 AGCCTAACTGGCGGGGACAAACCAACGAGTGAGGGTGAGGTCCAGATTGTGTCAACTGCT 120
QY 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyVala 60
DB 121 GCCCAAACTTCTCGCAACGTGCATCAATGGGTGTGCTGGACTGTCTACACGGGGCC 180
QY 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
DB 181 GGAACGAGGACCATCGCGTCACCCAAAGGTCTCTCATCCAGATGTATACCAATGTAGAC 240
QY 81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
DB 241 CAAGACCTTGTGGCTGGCGCGCTCCGCAAGGTAGCCGATCATTGACACCCCTGCACCT 300
QY 101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArq 120
DB 301 GGCTCTCGGACCTTACCTGTGTACAGGACGACGCGCGATGTCATTCCTGCGCGCGCGG 360
QY 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
DB 361 GGTGATAGCAGGGGAGCGCTGTGTGCGCGCGCGCCATTTCTACTTGAAGGCTCTCTCG 420
QY 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
DB 421 GGGGGTCCGCTTGTGTGCGCGCGGACGCGCGTGGGCATATTTAGGCGCGCGGTGTGC 480
QY 161 ThrArgGlyValAlaLysAlaValaspPheIleProValGluSerLeuGluThrThrMet 180
DB 481 ACCCGTGGAGTGGCTAAGCGGTGGACTTTATCCCTGTGGAGAACCTTAGAGACACCACT 540

QY 181 ArgSer 182
DB 541 AGGTCC 546

RESULT 3
ABK15344
ID ABK15344 standard; DNA: 2058 BP.
XX AC ABK15344;
XX AC ABK15344;
DT 08-MAY-2002 (first entry)
XX
XX Hepatitis C virus NS3/4a conformational epitope gene sequence.
XX
XX Hepatitis C virus; HCV; NS3/4a conformational epitope; seroconversion;
KW Immunoassay solid support; multiple epitope fusion antigen; MEFA;
KW non-structural protein; gene: ds.
XX
XX Hepatitis C virus.
OS
XX Key Location/Qualifiers
FH 1..2058
FT CDS /*tag- a
FT /partial
FT /product- "HCV NS3/4a conformational epitope"
FT /note- "This sequence lacks a stop codon"
XX

```

```

PN WO200196870-A2.
XX
PD 20-DEC-2001.
XX
PF 14-JUN-2001; 2001WO-US19156.
XX
PR 15-JUN-2000; 2000US-212082P.
PR 02-APR-2001; 2001US-280811P.
PR 02-APR-2001; 2001US-280867P.
XX
XX (CHIR ) CHIRON CORP.
XX
XX Chien DY, Arcangel P, Tandeske L, George-nascimento C, Coit D;
PI Medina-selby A;
XX
DR WPI; 2002-090228/12.
DR P-PSDB; AAU76377.
XX
XX Immunoassay solid support, useful for detecting hepatitis C virus
PT infection in biological sample, comprises HCV NS3/4a conformational
PT epitope and multiple epitope fusion antigen bound to the support .
XX
XX Disclosure; Fig 3; 92pp; English.
XX
XX The present invention relates to a new immunoassay solid support
XX consisting essentially of at least one hepatitis C virus (HCV) NS3/4a
XX conformational epitope and a multiple epitope fusion antigen (MEFA),
XX bound to the support. The NS3/4a conformational epitope and/or
XX MEFA reacts specifically with anti-HCV antibodies present in a biological
XX sample from an HCV-infected individual. The immunoassay of the invention
XX is useful for detecting hepatitis C virus infection in a biological
XX sample. The method of the invention provides a sensitive, accurate
XX diagnostic and prognostic tool to provide adequate patient care and to
XX prevent transmission of HCV by blood and by blood products, or by
XX personal contact. Use of NS3/4a conformational epitope in combination
XX with MEFA, provides a sensitive and reliable method for detecting early
XX HCV seroconversion. Use of MEFA has the added advantages of decreasing
XX masking problems, improving sensitivity in detecting antibodies by
XX allowing a greater number of epitopes on a unit surface area of
XX substrate, and improving substrate. Detection accuracy is increased and
XX the incidence of false results is reduced because of the identification
XX and the use of highly immunogenic HCV antigens which are present during
XX the early stages of HCV seroconversion. The present nucleic acid sequence
XX encodes the non-structural protein NS3/4a conformational epitope of the
XX invention.
XX
SQ Sequence 2058 BP; 419 A; 633 C; 581 G; 425 T; 0 other;

Alignment Scores:
Pred. No.:      1 46e-78      Length:      2058
Score:          946.00      Matches:      180
Percent Similarity: 100.00%      Conservative: 2
Best Local Similarity: 98.90%      Mismatches: 0
Query Match:      99.27%      Indels: 0
DB:              24      Gaps: 0

US-09-965-594-1 (1-182) x ABK15344 (1-2058)

QY 1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
DB 1 ATGGCGCCCATCACGGCGTAGCCGAGCAGACAAAGGGCCCTCTAGGTGCAATAATCAC 60
QY 21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
DB 61 AGCCTAACTGGCGGGGACAAACCAACGAGTGAGGGTGAGGTCCAGATTGTGTCAACTGCT 120
QY 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyVala 60
DB 121 GCCCAAACTTCTCGCAACGTGCATCAATGGGTGTGCTGGACTGTCTACACGGGGCC 180
QY 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
DB 181 GGAACGAGGACCATCGCGTCACCCAAAGGTCTCTCATCCAGATGTATACCAATGTAGAC 240
QY 81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
DB 241 CAAGACCTTGTGGCTGGCGCGCTCCGCAAGGTAGCCGATCATTGACACCCCTGCACCT 300
QY 101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArq 120
DB 301 GGCTCTCGGACCTTACCTGTGTACAGGACGACGCGCGATGTCATTCCTGCGCGCGCGG 360
QY 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
DB 361 GGTGATAGCAGGGGAGCGCTGTGTGCGCGCGCGCCATTTCTACTTGAAGGCTCTCTCG 420
QY 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
DB 421 GGGGGTCCGCTTGTGTGCGCGCGGACGCGCGTGGGCATATTTAGGCGCGCGGTGTGC 480
QY 161 ThrArgGlyValAlaLysAlaValaspPheIleProValGluSerLeuGluThrThrMet 180
DB 481 ACCCGTGGAGTGGCTAAGCGGTGGACTTTATCCCTGTGGAGAACCTTAGAGACACCACT 540

QY 181 ArgSer 182
DB 541 AGGTCC 546

RESULT 3
ABK15344
ID ABK15344 standard; DNA: 2058 BP.
XX AC ABK15344;
XX AC ABK15344;
DT 08-MAY-2002 (first entry)
XX
XX Hepatitis C virus NS3/4a conformational epitope gene sequence.
XX
XX Hepatitis C virus; HCV; NS3/4a conformational epitope; seroconversion;
KW Immunoassay solid support; multiple epitope fusion antigen; MEFA;
KW non-structural protein; gene: ds.
XX
XX Hepatitis C virus.
OS
XX Key Location/Qualifiers
FH 1..2058
FT CDS /*tag- a
FT /partial
FT /product- "HCV NS3/4a conformational epitope"
FT /note- "This sequence lacks a stop codon"
XX

```

Qy 81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
 Db :::
 241 CAACACCTTGCGGCTGGCCGCTCCGCAAGGTAGCCGATCATTGACACCTGCCTTGC 300
 Qy 101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 120
 Db GGTCTCTCGGACCTTTACCTGGTCACAGGACGCGCATGTCTCCCTGCGCGCGG 360
 Qy 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 140
 Db GGTGATAGCAGGCGGACCTGCTGCGCGCGCGCATTTCTACTTGAAGGCTCCTCG 420
 Qy 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaValCys 160
 Db GGGGCTCCGCTGTGTGCCCCCGGGCAGCGCTGGCATATTTAGGCGCGGTGTGC 480
 Qy 161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrMet 180
 Db ACCCGTGGAGTGGCTAAGGCGGTGGACTTATCCCTGTGGAGAACCTAGACACCATG 540
 Qy 181 ArgSer 182
 Db ::::::::::
 541 AGGTCC 546

RESULT 4
 ABX14410
 ID ABX14410 standard; DNA; 2058 BP.
 XX
 AC ABX14410;
 XX
 DT 06-MAR-2003 (first entry)
 XX
 DE DNA encoding HCV-1 NS3/4a conformational antigen.
 KW Immunoassay solid support; Hepatitis C virus type-1; HCV-1;
 KW NS3/4a conformational epitope; multiple epitope fusion antigen;
 KW HEPA; anti-HCV antibody; NS3/4a conformational antigen;
 KW HCV infection; mutant; gene; ds.
 XX
 OS Hepatitis C virus type 1.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT CDS 1..2058
 FT /*tag= a
 FT /partial
 FT /product= "NS3/4a conformational antigen"
 FT /note= "This sequence lacks a stop codon"
 XX
 PN US2002146685-A1.
 XX
 PD 10-OCT-2002.
 XX
 PF 14-JUN-2001; 2001US-0881654.
 XX
 PR 15-JUN-2000; 2000US-212082P.
 PR 02-APR-2001; 2001US-280811P.
 PR 02-APR-2001; 2001US-280867P.
 XX
 XX (CHIE/) CHIEN D Y.
 PA (ARCA/) ARCANDEL P.
 PA (TAND/) TANDESKE L.
 PA (GEOR/) GEORGE-NASCIMENTO C.
 PA (COIT/) COIT D.
 PA (MEDI/) MEDINA-SELBY A.
 XX
 PI Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Coit D;
 PI Medina-Selby A;
 XX
 DR WPI: 2003-147573/14.
 DR P-PSDB; ABG72261.
 XX

PT Immunoassay solid support for detecting Hepatitis C Virus infection in
 PT biological samples, comprises Hepatitis C Virus conformational epitope
 PT and multiple epitope fusion antigen -
 XX Disclosure; Fig 3A-3D; 45pp; English.

XX The present invention relates to immunoassays comprising Hepatitis C
 CC Virus (HCV) NS3/4a conformational epitope and multiple epitope fusion
 CC antigen (HEPA), bound to a solid support. The NS3/4a epitope and/or
 CC the multiple epitope fusion antigen react with anti-HCV antibodies
 CC present in a biological sample from an HCV-infected individual. The
 CC immunoassays and methods of the invention are useful for detecting
 CC HCV infection in a biological sample. The inventive immunoassay solid
 CC support provides a sensitive and reliable method for detecting early
 CC HCV seroconversion. The assays can detect HCV infection caused by any
 CC six known genotypes of HCV. The use of the multiple epitope fusion
 CC proteins decreases masking problems, improves sensitivity in detecting
 CC antibodies by allowing a greater number of epitopes on a unit area
 CC of substrate, and improves selectivity. The present sequence
 CC encodes HCV type 1 (HCV-1) NS3/4a conformational antigen, a mutant of
 CC the HCV-1 NS3/4a polypeptide.

SQ Sequence 2058 BP; 419 A; 633 C; 581 G; 425 T; 0 other;

Alignment Scores:

Pred. No.: 1.46e-78 Length: 2058
 Score: 946.00 Matches: 180
 Percent Similarity: 100.00% Conservative: 2
 Best Local Similarity: 98.90% Mismatches: 0
 Query Match: 99.27% Indels: 0
 DB: 25 Gaps: 0

US-09-965-594-1 (1-182) x ABX14410 (1-2058)

Qy 1 MetalProIleThrAlaTyrAlaGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
 Db 1 ATGGCGCCCATCAGCGGTACGCCAGCAGACAGAGGGGCTCTCTAGGTGCTAATACCC 60
 Qy 21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
 Db 61 AGCCTAACTGCGCGGGACAAAACCAAGTGGAGGTGAGTCCAGATTGTGTCACTGCT 120
 Qy 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
 Db 121 GCCCAACCTTCTCGCAACGTGCATCAATGGGTGTGCTGGACTGTCTACACGCGGCG 180
 Qy 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
 Db 181 GGAACGAGGACCATCGCTCACCAAGGTCCTGTCTATCCAGATGTATACCAATGTAGAC 240
 Qy 81 LysAspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
 Db 241 CAAGACCTTGCGGCTGGCCGCTCCGCAAGGTAGCCGATCATTGACACCTGCCTTGC 300
 Qy 101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 120
 Db 301 GGCTCTCGGACCTTTACCTGGTCACAGGACGCGCATGTCTCCCTGCGCGCGGCG 360
 Qy 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 140
 Db 361 GGTGATAGCAGGCGGACCTGCTGCGCGCGCGCATTTCTACTTGAAGGCTCCTCG 420
 Qy 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaValCys 160
 Db 421 GGGGCTCCGCTGTGTGCCCCCGGGCAGCGCTGGCATATTTAGGCGCGGTGTGC 480
 Qy 161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrMet 180
 Db 481 ACCCGTGGAGTGGCTAAGGCGGTGGACTTATCCCTGTGGAGAACCTAGACACCATG 540
 Qy 181 ArgSer 182
 Db ::::::::::
 541 AGGTCC 546

Key	Location/Qualifiers
FD	3..5360
FT	/*tag- a
FT	
XX	
PN	GB212511-A.
XX	
PD	26-JUL-1989.
XX	
XX	18-NOV-1988: 88GB-0027024.
XX	
XX	18-NOV-1987: 87US-0122714.
PR	30-DEC-1987: 87US-0139886.
PR	26-FEB-1988: 88US-0161072.
PR	26-OCT-1988: 88US-0263584.
XX	
XX	(CHIR) CHIRON CORPORATION.
PA	
XX	Houghton M, Choo QL, Kuo G;
XX	WPI: 1989-215054/30.
XX	
DR	
XX	
PT	Hepatitis C virus gene - used for prodn. of polynucleotide probes,
PT	polypeptide(s) and antibodies for diagnosis, prevention and treatment
PT	of infection.
XX	
PS	Disclosure; Fig. 26; 174pp; English.
XX	
CC	The sequence shows the composite cDNA sequence derived from the aligned
CC	hepatitis C virus (HCV) cDNA's in clones 14i, 11b, 7f, 7e, 8h, 33c, 40b,
CC	37b, 35, 36, 81, 32, 33b, 25c, 14c, 8f, 33f, 33g and 39c. The cDNA, 40b,
CC	encodes antigens which react with antibodies in patients with non-A
CC	non-B hepatitis (NANBH). The cDNA can be used to design probes, or to
CC	synthesize polypeptides, which are used to diagnose HCV-induced, or to
CC	to raise antibodies for immunoassay or treatment, or to produce
CC	vaccines. See also AAP90158, AAN90303-26, and AAN90328-36.
CC	(Updated on 25-MAR-2003 to correct PR field.)
XX	
SQ	Sequence 5360 BP; 1060 A; 1622 C; 1532 G; 1145 T; 1 other;
Alignment Scores:	
Pred. No.:	9.03e-78
Score:	943.00
Percent Similarity:	100.00%
Best Local Similarity:	98.35%
Query Match:	98.95%
DB:	10
US-09-965-594-1 (1-182) x AAN90327 (1-5360)	
QY	1 MetAlaProfilThrAlaTyAlaGlnGlnThrArgGlyLeuGlyCysIleIleThr 20
Db	CTGGCGCCATCAGCGGTACGCCAGCAGACAAAGGGGCTCTCTAGGTGCATAATCACC 989
QY	21 SerLeuThrClyArgAspLysAsnGlnValGluClyGluValGlnIleValSerThrAla 40
Db	AGCCTAACTGGCGGGGCAAAACCAAGTGAGGGTGAGGTCCAGATGTGTGCAACTGCT 1049
QY	41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrlHisGlyAla 60
Db	GCCCAAACTTCCTGGCAACGTCATCAATGGGGTGTGTGGACTGTCTACCAACGGGGCC 1109
QY	61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrlThrAsnValAsp 80
Db	GGAACGAGGACCATCGGTCCCAAGGGTCCGTATCCAGATGATATACCAATGTAGAC 1169
QY	81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuthrProCysThrCys 100
Db	CAAGACCTTCTGGCTGGCCCGCTCCGCCAAGGTAGCGGCTCATTGACACCC:GCACCTGC 1229
QY	101 GlySerSerAspLeuTyrlLeuValThrArgHisAlaAspValIleProValArgArgArq 120
Db	GGCTCTCGGACCTTACCTGGTCACGAGGACGCCGATCTCAATTCCTGTCGCCGCGCG 1289

Alignment Scores:

```

Pred. No.: 1,23e-77 Length: 6905
Score: 943.00 Matches: 179
Percent Similarity: 100.00% Conservatives: 3
Best Local Similarity: 98.35% Mismatches: 0
Query Match: 98.95% Indels: 0
DB: 10 Gaps: 0

US-09-965-594-1 (1-182) x AAN92103 (1-6905)

QY 1 MetalProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
DB 1203 CTGGGGCCCATCACGGCTACGCCAGCAGACAGAGGCCCTCTAGGGTGCATATACACC 1262

QY 21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
DB 1263 AGCCTTAACCTGGCGGGGACAAACCAACAGTGAGGGTGAGTCCAGATTGTCAACTGCT 1322

QY 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
DB 1323 GCCCAACCTTCTGGCAACGTGCATCAATGGGGTGTCTGGACTGTCTACACGGGGCC 1382

QY 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
DB 1383 GGAACGAGACCATCGCGTCACCCAGGGTCTGTCTATCCACAGATGATACCAATGTAGAC 1442

QY 81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
DB 1443 CAAGACCTTGTGGCTGGCCGCTCCGCAAGGTAGCGCTCATTCACACCCCTGCACCTGC 1502

QY 101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg 120
DB 1503 GGCTCTCTGGACCTTTACTGTGTACAGAGCGACGCGATGTCATTCCTGGCGCGGGCG 1562

QY 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
DB 1563 GGTGATAGCAGGGGAGCGCTGTCTGGCCCGGCCCATTCCTACTTGAAGGCTCTCTCG 1622

QY 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
DB 1623 GGGGGTCCGCTGTGTGGCCCGGGGCGACGCGGTGGGCATATTTAGGGCGCGGTGTGC 1682

QY 161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuCluThrThrMet 180
DB 1683 ACCCGTGGAGTGGCTAAGCGGTGGACTTTATCCCTGTGGAGACCTAGAGACCAACCATG 1742

QY 181 ArgSer 182
DB 1743 AGTCC 1748

RESULT 8
AAN92106
ID AAN92106 standard: DNA: 7310 BP.
AC
XX
AC AAN92106:
XX
XX
DT 25-MAR-2003 (updated)
XX
DT 02-MAR-1990 (first entry)
XX
DE Combined open reading frames of the hepatitis C virus (HCV) cDNAs from
DE clones K9-1 through 15e.
XX
XX Hepatitis C virus: HCV; non-A, non-B hepatitis; NANBH.
XX
OS Hepatitis C virus.
XX
XX Key location/Qualifiers
XX CDS 3..7310
XX /*tag= a
XX
XX EP318216-A.
XX
XX
XX
PD 31-MAY-1989.

```

```

XX 18-NOV-1988; 88EP-0310922..
XX
XX 18-NOV-1987; 87US-0122714.
PR 30-DEC-1987; 87US-0139886.
PR 26-FEB-1988; 88US-0161072.
PR 06-MAY-1988; 88US-0191263.
PR 26-OCT-1988; 88US-0263584.
PR 14-NOV-1988; 88US-0271450.
XX
XX (CHIR ) CHIRON CORP.
XX
XX Houghton M, Choo QL, Kuo G;
PI
XX
XX WPI: 1989-159274/22.
DR P-PSDB; AAP92050.
XX
XX Purified hepatitis C virus
PT - and associated nucleic acids and polypeptide(s)
XX
XX Claim 3: Figure 47-1 - 47-8; 139pp; English.
XX
XX It is a double-stranded nucleotide sequence of the open reading frame
CC (HCV) (tag a) extending through clones K9-1 to 15e of hepatitis C virus
CC (HCV) cDNA. It can be used to make oligomeric DNA hybridisation probes to
CC detect the presence of HCV nucleic acids in samples. The polypeptide(s)
CC it encodes could be used as immunoassay reagents and vaccines and to
CC generate antibodies useful in diagnosis and passive immunotherapy for
CC HCV infection/non-A, non-B hepatitis.
CC (Updated on 25-MAR-2003 to correct PR field.)
CC (Updated on 25-MAR-2003 to correct PI field.)
XX
XX Sequence 7310 BP; 1491 A; 2217 C; 2058 G; 1540 T; 4 other;

Alignment Scores:
Pred. No.: 1,32e-77 Length: 7310
Score: 943.00 Matches: 179
Percent Similarity: 100.00% Conservatives: 3
Best Local Similarity: 98.35% Mismatches: 0
Query Match: 98.95% Indels: 0
DB: 10 Gaps: 0

US-09-965-594-1 (1-182) x AAN92106 (1-7310)

QY 1 MetalProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
DB 1728 CTGGGGCCCATCACGGCTACGCCAGCAGACAGAGGCCCTCTAGGGTGCATATACACC 1787

QY 21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
DB 1788 AGCCTTAACCTGGCGGGGACAAACCAAGTGGAGGTGAGTCCAGATTGTCAACTGCT 1847

QY 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
DB 1848 GCCCAACCTTCTGGCAACGTGCATCAATGGGGTGTCTGGACTGTCTACACCGGGCC 1907

QY 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
DB 1908 GGAACGAGGACCATCGCGTCACCCAGGGTCTGTCTACAGATGTATACCAATGTAGAC 1967

QY 81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
DB 1968 CAAGACCTTGTGGCTGGCCCGCTCCGCAAGGTAGCGCTCATTCACACCGCTGCATTGC 2027

QY 101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg 120
DB 2028 GGCTCTCGGACCTTTACTGTGTACAGGACGCGGATGTCTATCCCGTGGCGGGCGG 2087

QY 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
DB 2088 GGTGATAGCAGGGGACGCTGTGTGGCCCGGCCCATTTCTCTACTTGAAGGCTCTCTCG 2147

QY 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValIleValPheArgAlaAlaValCys 160

```

```

Db 2148 GGGGTCCTGTGTGCCCCGGGACAGCGGTGGGCAIATTAGGGCCGGGTGTC 2207
Qy 161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
Db 2208 ACCGGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACACCATG 2267
Qy 181 ArgSer 182
Db 2268 AGGTCC 2273

RESULT 9
ID AAN90336 standard; DNA: 7310 BP.
XX AAN90336;
XX 25-MAR-2003 (updated)
DT 19-JUL-2001 (updated)
DT 01-NOV-1989 (first entry)
XX Composite hepatitis C virus (HCV) cDNA.
XX Hepatitis C virus; cDNA; clone 15e; clone k9-1; probe; vaccine; ds.
XX Pan troglodytes.
XX GB2212511-A.
PD 26-JUL-1989.
XX 18-NOV-1988; 88GB-0027024.
XX 18-NOV-1987; 87US-0122714.
PR 30-DEC-1987; 87US-0139886.
PR 26-FEB-1988; 88US-0161072.
PR 26-OCT-1988; 88US-0263584.
XX (CHIR) CHIRON CORPORATION.
XX Houghton M, Choo QL, Kuo G;
XX WPI: 1989-215054/30.
DR P-PSDB; AAP90288.
XX Hepatitis C virus gene - used for prodn. of polynucleotide probes,
PT polypeptide(s) and antibodies for diagnosis, prevention and treatment
PT of infection.
XX Disclosure; fig 47; 235pp; English.
XX The sequence shows a composite hepatitis C virus (HCV) cDNA, derived by
CC aligning clones k9-1 through 15e in 5'-3' direction. The cDNA
CC encodes antigens which react with antibodies in patients with non-A
CC non-B hepatitis (NANBH). The cDNA can be used to design probes, or to
CC synthesise polypeptides, which are used to diagnose HCV-induced NANBH.
CC to raise antibodies for immunoassay or treatment, or to produce
CC vaccines. See also AAP90288, and AAN90303-35.
CC (N.B. This record was resubmitted to correct errors in the sequence.)
CC (Updated on 25-MAR-2003 to correct PR field.)
XX
SQ Sequence 7310 BP; 1495 A; 2218 C; 2058 G; 1539 T; 0 other;

Alignment Scores:
Pred. No.: 1,32e-77 Length: 7310
Score: 943.00 Matches: 179
Percent Similarity: 100.00% Conservative: 3
Best Local Similarity: 98.35% Mismatches: 0
Query Match: 98.95% Indels: 0
DB: 10 Gaps: 0

US-09-965-594-1 (1-182) x AAN90336 (1-7310)

```

```

Cy 1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
1728 CTGGCCCAATACAGGGTACGCCACAGCAGACAGAGGGGCTCTCTAGGGTGCATATACCC 1787
Qy 21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
Db 1788 AGCTAACTGGCGGGACAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGCAACTGCT 1847
Qy 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
Db 1848 GCCCAACCTTCTGGCAACGTGCATCAATGGGGTGTGTGGACTGTCTTACCACGGGGCC 1907
Qy 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
Db 1908 GGAACGAGGACCATCGCGTCACCAAGGGTCTGTCTCATCCAGATGTATACCAATGTAGAC 1967
Qy 81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
Db 1968 CAAGACCTTGTGGCTGGCCCGCTCGCAAGGTAGCCGCTCATTTGACACCCCTGCATTGC 2027
Qy 101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg 120
Db 2028 GGCTCTCGGACCTTTACCTGGTCCAGGACGACGCCGATGTCAATCCCGCGCGCGCG 2087
Qy 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
Db 2088 GGTGATAGCAGGGGACGCTGCTGCGCCCGGCCCATTTCTCTACTTGAAGGCTCCTCG 2147
Qy 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
Db 2148 GGGGTGCTGTGTGCCCCGGGACGCGGTGGGCATATTTAGGCGCGGGTGTGC 2207
Qy 161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
Db 2208 ACCGGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACACCATG 2267
Qy 181 ArgSer 182
Db 2268 AGGTCC 2273

RESULT 10
AAQ98221
ID AAQ98221 standard; cDNA to mRNA; 7310 BP.
XX AAQ98221;
AC AAQ98221;
XX 25-MAR-2003 (updated)
DT 15-AUG-1996 (first entry)
XX Hepatitis C virus clone genome.
DE
XX Hepatitis C virus; HCV; antigen; detection; diagnosis; vaccine;
KW antibodies; immunoprophylaxis; sera; serum; ds.
XX Hepatitis C virus.
OS
XX US5443965-A.
PN
XX 22-AUG-1995.
PD
PF 05-APR-1991; 91US-0681703.
XX
XX 06-APR-1990; 90US-0505611.
PR
PR 09-OCT-1990; 90US-0594854.
XX (GENE-) GENELABS INC.
PA
XX Kim JP, Moockli R, Reyes GR;
PI
XX WPI: 1995-302120/39.
DR
XX New nucleic acids encoding hepatitis C virus antigens - used to
PT develop prods. for detection of HCV-infected sera and prodn. of
PT

```

PT vaccines and anti-HCV antibodies.

PS Example 4; Figure 11; 71pp; English.

XX Hepatitis C virus (HCV) antigens can be used for detecting HCV
 CC infected sera and individuals infected with HCV. They can also be
 CC used in an anti-HCV vaccine or for the production of anti-HCV
 CC antibodies which can be used for passive immunoprophylaxis. The
 CC antigens consistently identify more HCV positive serum samples with
 CC a high degree of specificity. See AAQ98202-14 and AAR81939-51.
 CC (Updated on 25-MAR-2003 to correct PF field.)
 CC (Updated on 25-MAR-2003 to correct PR field.)

XX SQ Sequence 7310 BP; 1494 A; 2217 C; 2060 G; 1539 T; 0 other;

Alignment Scores:

Pred. No.: 1.32e-77 Length: 7310
 Score: 943.00 Matches: 179
 Percent Similarity: 100.00% Conservatives: 3
 Best Local Similarity: 98.35% Mismatches: 0
 Query Match: 98.95% Indels: 0
 DB: 16 Gaps: 0

US-09-965-594-1 (1-182) x AAQ98221 (1-7310)

Qy 1 MetAlap-olietrAlaYrAlaGlnThrArgGlyLeuGlyCysIleIleThr 20
 Db 1728 CTGGCGCCCATCAGCGGTAGCCAGCAGCAGAGGGGCTCTAGGGTGCATATACACC 1787
 Qy 21 SerLeuThrGlyArgAspAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
 Db 1788 AGCCTAACTGGCGGCACAAACCAAGTGGAGGTGAGTCCAGATGTGTCAACATGCT 1847
 Qy 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
 Db 1848 GCCCAAACTTCCTGGCAACGTGCATCAATGGGTGTCTGAGCTGTCTACCCAGGGGCC 1907
 Qy 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetIYrThrAsnValAsp 80
 Db 1908 GGAACGAGGACCATCGCTCACCAAGGTCTCTCATCCAGATGTATACCAATGTAGAC 1967
 Qy 81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
 Db 1969 CAAGACCTTGTGGCTGGCGCTCCGCAAGGTAGCCGCTCATTGACACCTCGACTGC 2027
 Qy 101 GlySerSerAspLeuThrValThrArgHisAlaAspValIleProValArgArg 120
 Db 2028 GGTCTCTGGACCTTACTGTCTACGAGGACGCGGATGTCATTCCGTGCGCGCGG 2087
 Qy 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
 Db 2088 GGTGATAGCAGGGGAGCTGTCTGCGCCGCGCCCATTTCTACTTGAAGGCTCTCTCG 2147
 Qy 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
 Db 2148 GGGGTGCTCCTGTGTGCCCCGGGGGACCGCTGGGCATATTAGGCGCGCGGTGTC 2207
 Qy 161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
 Db 2208 ACCGTGGAGTGGCTAAGCGGTGACCTTATCCCTGTGGAGAACCTTAGACACCAACCA 2267
 Qy 181 ArgSer 182
 Db 2268 AGGTCC 2273
 RESULT 11
 ID AAA75296 standard; cDNA; 8316 BP.
 AC AAA75296;
 XX
 XX
 DT 15-JAN-2001 (first entry)
 XX

DE cDNA sequence compiled Hepatitis C virus cDNA clones.

XX Hepatitis C virus; HCV; antisense polynucleotide; polyprotein;
 KW viral infectivity; viral replication; ds.

XX Hepatitis C virus.

XX Key Location/Qualifiers
 CDS 1..8316
 /tag= a
 /note= "partial sequence; no termination codon given"

XX EP1034785-A2.

XX 13-SEP-2000.

XX 16-MAR-1990; 2000EP-0109602.

XX 17-MAR-1989; 89US-0325338.

XX 20-APR-1989; 89US-0341334.

XX 18-MAY-1989; 89US-0355002.

XX 16-MAR-1990; 90EP-0302866.

XX (CHIR) CHIRON CORP.

XX Houghton M, Choo Q, Kuo G;

XX WPI: 2000-566892/53.

XX P-PSDB; AAB18540.

XX Novel composition comprising a hepatitis C virus antisense
 PT polynucleotide which is complementary to or corresponds to a sense
 PT strand of the virus genome, and selectively hybridizes to it -

XX Example; Fig 16; 75pp; English.

XX The specification describes a pharmaceutical composition which
 CC comprises a hepatitis C virus (HCV) antisense polynucleotide. The
 CC HCV is characterized by a positive stranded RNA genome which has
 CC 40% homology at the polypeptide level to a HCV polyprotein. The
 CC antisense polynucleotide binds to cellular polynucleotides which
 CC enhance and/or are required for viral infectivity, replicative
 CC ability or chronicity. The antisense polynucleotides may also be
 CC designed to bind with high specificity, to be of increased stability,
 CC to be stable and to have low toxicity. The composition also comprises
 CC an agent which causes viral RNA to be inactive. The composition
 CC is used for preventing HCV replication in a system. The present
 CC sequence represents a novel HCV cDNA sequence, which is used in the
 CC course of the invention.

XX SQ Sequence 8316 BP; 1671 A; 2529 C; 2345 G; 1771 T; 0 other;

Alignment Scores:

Pred. No.: 1.55e-77 Length: 8316
 Score: 943.00 Matches: 179
 Percent Similarity: 100.00% Conservatives: 3
 Best Local Similarity: 98.35% Mismatches: 0
 Query Match: 98.95% Indels: 0
 DB: 21 Gaps: 0

US-09-965-594-1 (1-182) x AAA75296 (1-8316)

Qy 1 MetAlap-olietrAlaYrAlaGlnThrArgGlyLeuGlyCysIleIleThr 20
 Db 2734 CTGGCGCCCATCAGCGGTAGCCAGCAGCAGAGGGGCTCTAGGGTGCATATACACC 2793
 Qy 21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
 Db 2794 AGCCTAACTGGCGGCAGCAAAACCAAGTGGAGGTGAGTCCAGATGTGTCAACATGCT 2853
 Qy 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
 Db 2854 GCCCAAACTTCCTGGCAACGTGCATCAATGGGTGTCTGAGCTGTCTACCCAGGGGCC 2913

Qy 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
 Db 2914 GGAACGAGGACCATCGCTCCACCAAGGTCCTGTATCCACATGTATACCAATGTAGAC 2973
 Qy 81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
 Db 2974 CAAGACCTGTGGCTGGCCGCTCGGCAAGGTAGCCGCTCATTCACACCCCTGCATTCG 3033
 Qy 101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 120
 Db 3034 GGCTCTCGACCTTACCTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 3093
 Qy 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 140
 Db 3094 GGTGATAGGAGGCGACCTGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 3153
 Qy 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
 Db 3154 GGGGTGCGGTGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 3213
 Qy 161 ThrArgGlyValAlaLysAlaValAspPheIleProValCysSerLeuCluThrThrMet 180
 Db 3214 ACCCGTGGAGTGGCTAAGCGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 3273
 Qy 181 ArgSer 182
 Db 3274 AGGTCC 3279
 RESULT 12
 AAZ07656
 ID AAZ07656 standard; DNA; 9133 BP.
 XX
 AC AAZ07656;
 XX
 DT 20-MAR-2003 (updated)
 DT 08-NOV-1999 (first entry)
 XX
 DE Nucleotide sequence of HCV-1 ORF.
 XX
 KW Hepatitis C virus; HCV; J1; J7; HCV-1: non-A, non-B HCV; NANBH;
 KW HCV infection; vaccine; ds.
 XX
 OS Hepatitis C virus.
 XX
 FH Key Location/Qualifiers
 CDS 268..9132
 FT /*tag= b
 FT /trans= except- (pos:1588..1589; aa:Leu)
 FT /note= "this codon has an apparent 1 nucleotide deletion,
 which alters the reading frame"
 FT /trans= except- (pos:1647..1650; aa:Pro)
 FT /note= "this codon has an apparent 1 nucleotide
 insertion, which alters the reading frame; this
 insertion is not indicated in the sequence
 present in the format; sequence listing of the
 specification"
 XX
 PN EP939128-A2.
 XX
 XX 01-SEP-1999.
 XX
 XX 17-SEP-1990; 99FP-0101746.
 XX
 PR 15-SEP-1989; 89US-0408045.
 PR 21-DEC-1989; 89US-0456142.
 PR 17-SEP-1990; 90EP-0310149.
 XX
 PA (CHIR) CHIRON CORP.
 PA (OYAA/) OYA A.
 XX
 PI Cha T, Han J, Houghton M, Irvine BD, Kolberg JA;
 PI Miyamura T, Saito I, Weiner AJ;

XX WFI: 1999-480843/41.
 DR P-PSDB; AAY14975.
 XX
 PT New Hepatitis C Virus isolates, useful for diagnosis of hepatitis
 infections and development of vaccines
 XX
 PS Disclosure; Fig 12; 132pp; English.
 XX
 CC The invention provides two new isolates of hepatitis C virus (HCV), J1
 and J7. These two isolates comprise nucleotide and amino acid sequences
 CC that are distinct from the HCV isolate HCV-1. The nucleotide sequences
 CC may be used to detect non-A, non-B HCV (NANBH) polynucleotides by
 CC hybridisation for diagnosis of NANBH infections. They may also be used to
 CC screen blood donors, donated blood and blood products for this infection.
 CC The isolates may also be used to isolate other naturally occurring
 CC variants of the virus. The polypeptides may be used as a vaccine for
 CC administration to patients to protect against infection with NANBH. The
 CC present sequence represents the nucleotide sequence of HCV-1 ORF.
 CC (Updated on 20-MAR-2003 to correct PF field.)
 CC (Updated on 20-MAR-2003 to correct PR field.)
 XX
 SQ Sequence 9133 BP; 1834 A; 2772 C; 2600 G; 1927 T; 0 other;
 Alignment Scores:
 Pred. No.: 1.74e-77 Length: 9133
 Score: 943.00 Matches: 179
 Percent Similarity: 100.00% Conservative: 3
 Best Local Similarity: 98.35% Mismatches: 0
 Query Match: 98.95% Indels: 0
 DB: 20 Gaps: 0
 US-09-965-594-1 (1-182) x AAZ07656 (1-9133)
 Qy 1 MetaLaproteinHraLalaTyrAlaGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
 Db 3343 CTGGCCCATACGGGTACGGCCAGCAGCAGAGGGGCTCTAGGTGTCATAATCACC 3402
 Qy 21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
 Db 3403 AGCTAACTGCGCGGACAAAACCAAGTGGAGGTGAGTCCAGATTGTCTCAACTGCT 3462
 Qy 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
 Db 3463 GCCCAACCTCTCTGGCAACGTGCATCAATGGGTGTGTGGACTCTTACCACGGGGCC 3522
 Qy 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
 Db 3523 GGAACGAGGACCATCGCTCACCAAGGTCTGTCTATCCAGATGTATACCAATGTAGAC 3582
 Qy 81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
 Db 3583 CAAGACCTTGTGGCTGGCCGCTCGCAAGGTAGCGCTCATTCACACCCCTGCATTCG 3642
 Qy 101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 120
 Db 3643 GGCTCTCGACCTTACCTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 3702
 Qy 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 140
 Db 3703 GGTGATAGGAGGCGACCTGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 3762
 Qy 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
 Db 3763 GGGGTGCGGTGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 3822
 Qy 161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrMet 180
 Db 3823 ACCGTGGAGTGGCTAAGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACACCATG 3882
 Qy 181 ArgSer 182
 Db 3883 AGGTCC 3888

```

RESULT 13
AAQ05956
ID AAQ05956 standard; DNA: 9185 BP.
AC
XX AAQ05956;
XX
XX 25-MAR-2003 (updated)
DT 23-JAN-1991 (first entry)
XX
XX Sense strand of the compiled Hepatitis C virus cDNA sequence.
DE
XX Hepatitis C virus (HCV); antiviral agent: ss.
XX
XX Hepatitis C virus.
OS
XX
XX Key Location/Qualifiers
FH CDS 320..9185
FT FT /*tag= a
FT misc_RNA 1..1667
FT /*tag= b
FT /*note="epitope within this region is claimed"
FT misc_RNA 8978..9185
FT /*tag= c
FT /*note="encodes an epitope that is claimed"
XX
XX EP388232-A.
XX
XX 19-SEP-1990.
XX
XX 16-MAR-1990; 90EP-0302866.
XX
XX 17-MAY-1989; 89US-0355002.
XX 18-MAR-1989; 89US-0325338.
XX 20-APR-1989; 89US-0341334.
XX
XX (CHIR ) CHIRON CORP.
XX
XX Houghton M, Choo QL, Kuo G;
XX
XX WPI: 1990-284418/38.
XX P-PSDB; AAR08124.
XX
XX Hepatitis C virus DNA - used for producing probes,
XX polypeptide(s), antibodies and anti-sense polynucleotide(s) for
XX diagnosis and therapy.
XX
XX Disclosure; Fig 17; 83pp; English.
XX
XX HCV cDNA libraries were constructed using pooled serum from a
XX chimpanzee with chronic HCV infection. A lambda gtl1 library was
XX screened with probes derived from previously isolated clones. The
XX ORF is derived from the overlapping clones b114a, ag30a, CA205a,
XX CA290a, CA216a, p14a, CA167b, CA156e, CA84a, CA59a, K9-1, 26j, 13i,
XX 12f, 14i, 11b, 7f, 8h, 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c,
XX 14c, 8f, 33f, 33g, 39c, 35f, 19g, 26g, 15e, b5a and 16jh. These
XX clones extend the sequence of the HCV genome reported in Ep-318216.
XX The upstream region from nucleotides -319 to -1348 (-1-1667 in this
XX file) is covered by clones b114a, 18g, ag30a, CA205a, CA290a,
XX CA216a, p14a, CA167b, CA156e, CA84a and CA59a; nucleotides
XX 8659-8866 (-8978-9185 in this file) are covered by clones b5a and
XX 16jh.
XX
XX See also AAQ05955.
XX (Updated on 25-MAR-2003 to correct PA field.)
XX
XX Sequence 9185 BP; 1849 A; 2790 C; 2608 G; 1938 T; 0 other;

Alignment Scores:
Pred. No.: 1.75e-77 Length: 9185
Score: 943.00 Matches: 179
Percent Similarity: 100.00% Conservatives: 3
Best Local Similarity: 98.35% Mismatches: 0
Query Match: 98.95% Indels: 0

```

```

DB: 11 Gaps: 0
US-09-965-594-1 (1-182) x AAQ05956 (1-9185)
QY 1 MetAlaProIleThrAlaIleThrArgGlyLeuLeuGlyCysIleIleThr 20
Db 3395 CTGGCGCCCATCAGCGCTAGCCAGCAGACAAGGGCCCTCCTAGGTGCATAATCACC 3454
QY 21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
Db 3455 AGCCTAACCTGGCGGGGACAAAACCAAGTGGAGGTGAGGTCCAGATTGTGCAACTGCT 3514
QY 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValThrHisGlyAla 60
Db 3515 GCCCAAACTTCCTGGCAACGTGCATCAATGGGTGTGCTGGACTGTCTACACGGGGCC 3574
QY 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
Db 3575 GGAACGAGGACCATCGCGTCACCAAGGGTCTGTGCATCCAGATGTATACCAATGTAGAC 3634
QY 81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
Db 3635 CAGACCTTGTGGGCTGGCCCGCTCCGCAAGGTAGCCGCTCATTTGACACCTGACCTGC 3694
QY 101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg 120
Db 3695 GGCTCTCGGACCTTTACCTGTGTACGAGGACGCGCATGTCTCCGCTGCGCGCGCGG 3754
QY 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
Db 3755 GGTGATAGCAGGGGACGCTGTGCGCCCGCGCGGACGCGCTATTTAGGCGCGGTGTGC 3814
QY 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
Db 3815 GGGGTCCGCTGTGTGCGCCCGCGGGGACGCGCGTGGGCATATTTAGGCGCGGTGTGC 3874
QY 161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
Db 3875 ACCCGTGAGTGCTAAGCGGTGGACTTTA-CCCTGTGGAGAACCTAGAGAACCATG 3934
QY 181 ArgSer 182
Db 3935 AGGTCC 3940
RESULT 14
AAQ10566
ID AAQ10566 standard; DNA: 9185 BP.
XX
XX AAQ10566;
XX
XX 25-MAR-2003 (updated)
DT 29-APR-1991 (first entry)
XX
XX Hepatitis C virus strain 1 DNA.
XX
XX Hepatitis C virus; HCV-1; non-A, non-B hepatitis; HCV antigen;
XX viral infections; ss.
XX
XX Hepatitis C virus.
XX
XX EP414475-A.
XX
XX 27-FEB-1991.
XX
XX 21-AUG-1990; 90EP-0309120.
XX
XX 25-AUG-1989; 89US-0398667.
XX
XX (CHIR ) CHIRON CORP.
XX
XX Weiner AJ, Steimer KS;
XX WPI; 1991-059670/09.

```

XX Cell lines infected with hepatitis C virus - are used as source
 PT of antigens for detection of HCV antibodies, for vaccines, and
 PT for screening anti-viral agents
 XX
 XX Disclosure; fig 1: 24pp; English.
 XX This is a hepatitis C virus (HCV) composite cDNA sequence, deduced
 CC using overlapping clones, a compsn. contg. the antigenic protein
 CC encoded by this sequence is useful for detecting anti-HCV anti-
 CC bodies (Abs) and for screening an agent which inhibits HCV repli-
 CC cation. A cell line infected with this virus can be used as a
 CC source of antigens. The antigen is useful for preparing vaccines
 CC for treating viral infections. See also AAQ10567.
 CC (Updated on 25-MAR-2003 to correct PA field.)
 XX

SQ Sequence 9185 BP; 1849 A; 2790 C; 2608 G; 1938 T; 0 other;

Alignment Scores:
 Pred. No.: 1.75e-77 Length: 9185
 Score: 943.00 Matches: 179
 Percent Similarity: 100.00% Conservativeness: 3
 Best Local Similarity: 98.35% Mismatches: 0
 Query Match: 98.95% Indels: 0
 DB: 12 Gaps: 0

US-09-965-594-1 (1-182) x AAQ10566 (1-9185)

QY 1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuGlyCysIleIleThr 20
 DB :::
 3395 CTGGCGCCCATCAGCGGCTACGCCAGCAGACAGGGGCTCTAGGGTGCAATACACC 3454
 QY 21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyValGlnIleValSerThra 40
 DB :::
 3455 AGCCTAACTGCGCGGACAAAAACCAAGTGGAGGTGAGTCCAGATGTGTCAACTGCT 3514
 QY 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
 DB :::
 3515 GCCCAAAACCTTCCTGGCAACGTGCATCAATGGGGTGTGGACTGTCTACCAAGGGGCC 3574
 QY 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
 DB :::
 3575 GGAACGAGGACCATCGCTGCTACCCAGGGTCTCTGTCATCCAGATGTATACCAATGTAGAC 3634
 QY 81 LysAspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
 DB :::
 3635 CAAGACCTTGTGGGTGCGCCGCTCCGCAAGGTAGCGCTCATITGACACCCCTGCACCTGC 3694
 QY 101 GlySerSerAspLeuTyrLeuValThrArqHisAlaAspValIleProValArgArgArg 120
 DB :::
 3695 GGCTCTCGGACCTTACTCGTCACAGGACCGCGATGTCTATCCGTCGCGCGCGG 3754
 QY 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
 DB :::
 3755 GGTGATAGCAGGGCAGCGCTGCTGCGCCCGCGCCCATTTCTACTTGAAGGCTCTCTCG 3814
 QY 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
 DB :::
 3815 GGGGTGCGCTGTGTGCGCCCGGGGACCGCGGTGGGATATTTAGGCGCGGTGTGC 3874
 QY 161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
 DB :::
 3875 ACCCTGGAGTGGCTAAGCGGGTGGACTTTATCTCTGTGGAGAACCTAGACACACCATG 3934
 QY 181 ArgSer 182
 DB ::::::::::
 3935 AGGTCC 3940

RESULT 15

AA75297

ID AAA75297 standard: cDNA; 9185 BP.

XX

AC AAA75297;

XX 15-JAN-2001 (first entry)
 DI Sense strand of HCV encoding a polyprotein.
 DE
 XX
 XX Hepatitis C virus; HCV; antisense polynucleotide; polyprotein;
 KW viral infectivity; viral replication; ds.
 XX
 XX Hepatitis C virus.
 OS
 XX
 XX Key Location/Qualifiers
 CDS 320..9184
 FT /*tag= a
 FT /note= "partial sequence; no termination codon given"
 XX
 XX EP1034785-A2.
 XX
 XX 13-SEP-2000.
 PD
 XX
 XX 16-MAR-1990; 2000EP-0109602.
 PF
 XX
 XX 17-MAR-1989; 89US-0325338.
 PR
 XX 20-APR-1989; 89US-0341334.
 PR
 XX 18-MAY-1989; 89US-0355002.
 PR
 XX 16-MAR-1990; 90EP-0302866.
 PR
 XX (CHIR) CHIRON CORP.
 PA
 XX
 XX Houghton M, Choo Q, Kuo G;
 X1
 XX WPI: 2000-566891/53.
 DR P-PSDB; AAB18541.
 XX
 XX Novel composition comprising a hepatitis C virus antisense
 PT polynucleotide which is complementary to or corresponds to a sense
 PT strand of the virus genome, and selectively hybridises to it -
 XX
 XX Example; Fig 17: 75pp; English.
 XX
 XX The specification describes a pharmaceutical composition which
 CC comprises a hepatitis C virus (HCV) antisense polynucleotide. The
 CC HCV is characterized by a positive stranded RNA genome which has
 CC 40% homology at the polypeptide level to a HCV polyprotein. The
 CC antisense polynucleotide binds to cellular polynucleotides which
 CC enhance and/or are required for viral infectivity, replicative
 CC ability or chronicity. The antisense polynucleotides may also be
 CC designed to bind with high specificity, to be of increased stability,
 CC to be stable and to have low toxicity. The composition also comprises
 CC an agent which causes HCV RNA to be inactive. The composition
 CC is used for preventing HCV replication in a system. The present
 CC sequence represents a novel HCV cDNA sequence, which is used in the
 CC course of the invention.
 XX
 SQ Sequence 9185 BP; 1849 A; 2790 C; 2608 G; 1938 T; 0 other;

Alignment Scores:

Pred. No.: 1.75e-77 Length: 9185
 Score: 943.00 Matches: 179
 Percent Similarity: 100.00% Conservativeness: 3
 Best Local Similarity: 98.35% Mismatches: 0
 Query Match: 98.95% Indels: 0
 DB: 21 Gaps: 0

US-09-965-594-1 (1-182) x AAA75297 (1-9185)

QY 1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuGlyCysIleIleThr 20
 DB :::
 3395 CTGGCGCCCATCAGCGGCTACGCCAGCAGACAGGGGCTCTAGGGTGCAATACACC 3454
 QY 21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyValGlnIleValSerThra 40
 DB :::
 3455 AGCCTAACTGCGCGGACAAAAACCAAGTGGAGGTGAGTCCAGATGTGTCAACTGCT 3514

Qy	41	AlaGlnThrPheIcuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla	60
Dd	3515	GCCAAACCTTCCTGGCAAGCTGCATCAATGGGTGTGCTGGACTGCTACCACGGGGCG	3574
Qy	61	GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAspValAsp	80
Dd	3575	GGAACGAGGACCATCGCGTCACCAAGGTCTCTGTCATCCAGATGATACCAATGTAGAC	3634
Qy	81	LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys	100
Dd	3635	CAAGACCTTGTTGGCTGCCCGCTCCGCCAAGTAGCGCTCATTTGACACCCCTGCACCTTGC	3694
Qy	101	GlySerSerAspLeuTyrIleuValThrArgHisAlaAspValIleProValArgArgArg	120
Dd	3695	GGCTCTCGAGACCTTTACCTGGTCACGAGCACGCCGATGCAITCCGTGCCCGCGCGG	3754
Qy	121	GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer	140
Dd	3755	GGTGATAGCAGGGCACCTTGCTGCGCCCGGCCCATTTCTTACTTGAAGGCTCCTCG	3814
Qy	141	GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys	160
Dd	3815	GGGGGTCGGTGTGTGTCGCCCGGGGCACGCCGTGGGCATATTATTAGGGCGCGGGTGTGC	3874
Qy	161	ThrArgGlyValAlaIatysAlaValAspPheIleProValGluSerLeuGluThrThrMet	180
Dd	3875	ACCCGTGGATGGCTAAGCGGGTGGACTTTATCCCTGTGGAGAACCTTAGAGACAACCATG	3934
Qy	181	ArgSer	182
Dd	3935	AGGTCC	3940

Search completed: August 30, 2003, 19:47:38
Job time : 188.009 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 18:01:52 ; Search time 9.01384 Seconds
(without alignments)
949.524 Million cell updates/sec

Title: US-09-965-594-1

Perfect score: 953

Sequence: 1 MAPITAYAQTRGILGCIIT.....GVAKAVDFIPVESLFTTMS 182

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 segs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	943	99.0	3011	1	POLG_HCV1
2	927	97.3	3011	1	POLG_HCVH
3	892	93.6	3010	1	POLG_HCVTW
4	891	93.5	3010	1	POLG_HCVBK
5	887	93.1	3010	1	POLG_HCVJA
6	884	92.8	3010	1	POLG_HCVJT
7	714	74.9	3033	1	POLG_HCVJB
8	712	74.7	3033	1	POLG_HCVJ6
9	87	9.1	209	1	PAAD_PSEAE
10	84	8.8	321	1	HHQA_ARATH
11	82	8.6	452	1	AAMP_HUMAN
12	82	8.6	485	1	Y136_TREPA
13	80.5	8.4	437	1	DEGL_ARATH
14	75	7.9	253	1	CAC3_BOVIN
15	74.5	7.8	415	1	ZP3_RABIT
16	74.5	7.8	776	1	HYFE_AZOVI
17	74.5	7.8	911	1	TB11_NEIMB
18	74	7.8	326	1	PANE_RHIL0
19	73.5	7.7	263	1	GRAK_MOUSE
20	73	7.7	730	1	HEL3_METMA
21	72.5	7.6	257	1	GRAM_HUMAN
22	72	7.6	627	1	CAJH_MOUSE
23	72	7.6	1527	1	CAJH_MOUSE
24	72	7.6	2663	1	CENE_HUMAN
25	72	7.6	3491	1	ERV1_SACER
26	71.5	7.5	248	1	GRAD_MOUSE
27	71.5	7.5	323	1	VPRT_SMRVH
28	71	7.5	219	1	SPR1_IPOBA
29	71	7.5	336	1	ULL16_EBV
30	71	7.5	529	1	PGL2_RALSO
31	71	7.5	1180	1	ITAI_RAT
32	70.5	7.4	264	1	CTRL_HUMAN
33	70.5	7.4	659	1	VST2_HEVME

34	70.5	7.4	743	1	TPE3_HUMAN
35	70	7.3	280	1	VIE3_HCMVT
36	70	7.3	410	1	VIE2_HCMVT
37	70	7.3	478	1	MM03_RABIT
38	70	7.3	730	1	HEUS_METAC
39	69.5	7.3	915	1	TBPI_NEIGO
40	69	7.2	355	1	CWG2_SCHPO
41	69	7.2	637	1	DNAK_BRUME
42	69	7.2	1136	1	C4BA_BACTI
43	69	7.2	1180	1	C4AA_BACTI
44	69	7.2	4391	1	PGBM_HUMAN
45	68.5	7.2	255	1	CATG_HUMAN

ALIGNMENTS

RESULT 1
POLG_HCV1 STANDARD; PRT; 3011 AA.
AC P26664;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
DE Hepatitis C virus (isolate 1) (HCV).
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.
OC NCBI_TaxID=11104;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91172826; PubMed=1848704;
RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C., Gallegos C., Colt D., Medina-Selby A., Barr P.J., Weiner A.J., Bradley D.W., Kuo G., Houghton M.;
RA *Genetic organization and diversity of the hepatitis C virus.*;
RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION. NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate + (RNA)(N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M62321; AAA45676.1; -
CC PIR: A39166; GNMV03
CC PDB: 1AIV; 16-FEB-99.
CC PDB: 1HEI; 25-NOV-98.
CC MEROPS: S29.001; -
CC MEROPS: U39.001; -
CC InterPro: IPR001410; DEAD
CC InterPro: IPR002522; HCV_capsid.

DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRp.
 DR InterPro: IPR001690; Helicase_C.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRp; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXdc; 1.
 DR PolyProtein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 192 191 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 1007 1615 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4 (POTENTIAL).
 FT CHAIN 2014 3011 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 347 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT TRANSMEM 347 369 POTENTIAL.
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;
 Query Match 99.0%; Score 943; DB 1; Length 3011.
 Best Local Similarity 98.48; Pred. No. 3.4e-82.
 Matches 179; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MAPITAYAQOTRGLGCIITSLTGRDKNOVEGEVQIVSTAOTFLATCINGVCWTVYHGA 60

Db 1026 LAPITAYAQOTRGLGCIITSLTGRDKNOVEGEVQIVSTAOTFLATCINGVCWTVYHGA 1085
 Qy 61 GTRTIASPKGPVTOMYTNVDKDLVGNPAPOGSGRSITPCTCGSSDLYLTVRHADVTPVRRR 120
 Db 1086 GTRTIASPKGPVTOMYTNVDQDLVGNPAPOGSGRSITPCTCGSSDLYLTVRHADVTPVRRR 1145
 Qy 121 GDSRGLLSPPRISYLUKSGGGPLLCAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 180
 Db 1146 GDSRGLLSPPRISYLUKSGGGPLLCAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 1205
 Qy 181 RS 182
 Db 1206 RS 1207
 RESULT 2
 POLG_HCVH STANDARD; PRT: 3011 AA.
 AC P27958;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate H) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=11108;
 RN [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE-92052256; PubMed-1658800;
 RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
 RA Prince A.M.;
 RT "Genomic structure of the human prototype strain H of hepatitis C
 RT virus: comparison with American and Japanese isolates.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
 RP [2]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
 RX MEDLINE-97331322; PubMed-9187654;
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
 RT "Structure of the hepatitis C virus RNA helicase domain.";
 RL Nat. Struct. Biol. 4:463-467(1997).
 RP [3]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
 RX MEDLINE-98154321; PubMed-9493270;
 RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
 RA Murcko M.A., Lin C., Caron P.R.;
 RT "Hepatitis C virus NS3 RNA helicase domain with a bound
 RT oligonucleotide: the crystal structure provides insights into the mode
 RT of unwinding.";
 RL Structure 6:89-100(1998).
 CC [1-] FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
 CC [1-] FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
 CC [1-] FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
 CC ACTIVATION OF NS3.
 CC [1-] FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
 CC [1-] FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
 CC [1-] CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the p6
 CC position, Cys or Thr in p1 and Ser or Ala in p1'.
 CC [1-] CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC (RNA)(N).
 CC [1-] SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.

QY 1 MAPITAYAOOTGRLCCCTTSLTGKKNQVGEVQVWIAAQTFLATCINGCVTVYHGA 60
 DB 1026 LAPITAYAOOTGRLGCGCTTSLTGKKNQVGEVQVWIAAQTFLATCINGCVTVYHGA 1085
 QY 61 GTTITASPKGPVIOMYTNVDKLVGWPAPQGSRLTPTCTGSSDLYLVTRHADYIPVRRR 120
 DB 1086 GSKTIAGPKGPITOMYTNVDKLVGWPAPQGSRLTPTCTGSSDLYLVTRHADYIPVRRR 1145
 QY 121 GDSRGLSPRISYLVKSSGGPLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 180
 DB 1146 GDSRGLSPRISYLVKSSGGPLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESMETTM 1205
 QY 181 RS 182
 DB 1206 RS 1207

RESULT 4
 ID POLG_HCVBK STANDARD; PRT: 3010 AA.
 AC P26663;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48);
 OS Hepatitis C virus (isolate BK) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=11105;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91140698; PubMed=1847440;
 RA Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
 RA Onishi E., Andoh T., Yoshida I., Okayama H.;
 RT "Structure and organization of the hepatitis C virus genome isolated
 RT from human carriers.";
 RL J. Virol. 65:1105-1113(1991).
 RN [2]
 RP SEQUENCE OF 1487-1500.
 RX MEDLINE=96235224; PubMed=8647104;
 RA Borowski P., Heiland M., Oehlmann K., Becker B., Korneleky L.;
 RT "Non-structural protein 3 of hepatitis C virus inhibits
 RT phosphorylation mediated by cAMP-dependent protein kinase.";
 RL Eur. J. Biochem. 237:611-618(1996).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
 RX MEDLINE=97015088; PubMed=8861916;
 RA Love R.A., Parge H.E., Wickersham J.A., Hostomsky Z., Habuka N.,
 RA Moomaw E.W., Adachi I., Hostomsky Z.;
 RT "The crystal structure of hepatitis C virus NS3 proteinase reveals a
 RT trypsin-like fold and a structural zinc binding site.";
 RL Cell 87:331-342(1996).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
 RX MEDLINE=98227846; PubMed=9568891;
 RA Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
 RA Steinkuehler C., Tonel L., de Francesco R., Kuo L.C., Chen Z.;
 RT "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C
 RT virus: a 2.2-A resolution structure in a hexagonal crystal form.";
 RL Protein Sci. 7:837-847(1998).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +

(RNA)(N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M58335; AAA72945.1; -
 CC PIR: A38465; GHWYTC.
 CC PDB: 1A1Q; 25-MAR-98.
 CC PDB: 1JXP; 14-JAN-98.
 CC PDB: 1NS3; 08-APR-98.
 CC PDB: 1C2P; 15-NOV-00.
 CC PDB: 1CSJ; 08-NOV-99.
 CC PDB: 1GX5; 09-APR-02.
 CC PDB: 1GX6; 10-APR-02.
 CC PDB: 1QUV; 26-JUN-00.
 CC PDB: 8OHM; 20-APR-99.
 CC MEROPS: S29.001; -
 CC MEROPS: U39.001; -
 CC InterPro: IPR001410; DEAD.
 CC InterPro: IPR002522; HCV_capsid.
 CC InterPro: IPR002521; HCV_Core.
 CC InterPro: IPR002519; HCV_env.
 CC InterPro: IPR002531; HCV_NS1.
 CC InterPro: IPR002518; HCV_NS2.
 CC InterPro: IPR004109; HCV_NS3.
 CC InterPro: IPR000745; HCV_NS4a.
 CC InterPro: IPR001490; HCV_NS4b.
 CC InterPro: IPR002868; HCV_NS5a.
 CC InterPro: IPR002166; HCV_RdRP.
 CC InterPro: IPR007095; RNA_pol_DS_PS.
 CC InterPro: IPR007094; RNA_pol_PSVir.
 CC Pfam: PF01543; HCV_capsid; 1.
 CC Pfam: PF01542; HCV_core; 1.
 CC Pfam: PF01539; HCV_env; 1.
 CC Pfam: PF01560; HCV_NS1; 1.
 CC Pfam: PF01538; HCV_NS2; 1.
 CC Pfam: PF02907; HCV_NS3; 1.
 CC Pfam: PF01006; HCV_NS4a; 1.
 CC Pfam: PF01001; HCV_NS4b; 1.
 CC Pfam: PF01506; HCV_NS5a; 1.
 CC Pfam: PF00998; Viral_RdRP; 1.
 CC ProDom: PD186062; HCV_NS1; 1.
 CC SMART: SM00487; DEXDc; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CELLULAR AMINOPEPTIDASE.
 FT CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT CHAIN 347 369 POTENTIAL.
 FT ACT_SITE 1083 CHARGE RELAY SYSTEM.
 FT ACT_SITE 1107 CHARGE RELAY SYSTEM.
 FT ACT_SITE 1165 CHARGE RELAY SYSTEM.
 FT NP_BIND 1230 ATP (POTENTIAL).

	Qy	1	MAPTAYAAQTGRLLGCIITSLTGRDNKNOVEGEVQIVSTTAOFTIATCINGVCVTYHGA	60
	Dd	1026	LAPITAYSQQTRGLLGCIITSLTGRDNKNOVEGVQVWSTATQSFLATCVNGVCVTYHGA	1085
	Qy	61	GTRTIASPKGPVIOMYTNDVKDLVGMPAQSGRSLTPTCTCGSSDLYLVYTRHADYIPVRRR	120
	Dd	1086	GSKTLAAPPKGPIOTMYNTVDQLVGNPKPPGARSLTPCTCGSSDLYLVYTRHADYIPVRRR	1145
	Qy	121	GDSRGSLSPRPISYLXGSSGGPLLCPAGHAVICTFRAAVCTRGVAKAVDFIPVRSLETTM	18C
	Dd	1146	GDSRGSLSPRVSYLXGSSGGPLLCPFHAVIGTFRAAVCTRGVAKAVDEVVESMEITM	120S
	Qy	181 RS 182		
		I I		

1206 RS 1207

Db

RESULT 5

POLG_HCVJA

ID POLG_HCVJA STANDARD; PRT; 3010 AA.

AC P26662;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].

DE Hepatitis C virus [Isolate Japanese] (HCV). Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].

OS Hepatitis C virus [Isolate Japanese] (HCV). Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].

OC Hepacivirus.

OC NCBI_TaxID=11116;

RR [1]

RR SEQUENCE FROM N.A.

RR MEDLINE=91088550; PubMed=2175903;

RR Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S., Sugimura T., Shimotohno K.;

RA "Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis.,"

RI Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).

RL [2]

RR DISCUSSION OF SEQUENCE.

RR MEDLINE=91192160; PubMed=1849488;

RR Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraishi K., Ohkoshi S., Shimotohno K.;

RA "Molecular structure of the Japanese hepatitis C viral genome.,"

RL FEBS Lett. 280:325-328(1991).

CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION. NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.

CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.

CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate + [RNA](N).

CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPID PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.

CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on way use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

CC EMBL: D90208; BAAL4233.1; -.

DR DR HSR: A39253; GNMVQJ.

DR HSR: P26663; IJXP.

DR MEROPS: S29.001; -.

DR MEROPS: U39.001; -.

DR InterPro: IPR001410; DEAD.

DR InterPro: IPR002522; HCV_capsid.

DR InterPro: IPR002521; HCV_core.

DR InterPro: IPR002519; HCV_env.

DR InterPro: IPR002531; HCV_NS1.

DR InterPro: IPR002538; HCV_NS2.

DR InterPro: IPR004109; HCV_NS3.

DR InterPro: IPR000745; HCV_NS4A.

DR InterPro: IPR001490; HCV_NS4B.

DR InterPro: IPR002868; HCV_NS5A.

DR Pfam: PF02907: HCV_NS3: 1.
 DR Pfam: PF01006: HCV_NS4a: 1.
 DR Pfam: PF01001: HCV_NS4b: 1.
 DR Pfam: PF01506: HCV_NS5a: 1.
 DR Pfam: PF00271: helicase_C: 1.
 DR Pfam: PF00998: Viral_RdRP: 1.
 DR ProDom: PD186062: HCV_NS1: 1.
 DR SMART: SM00487: DEXDC: 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CARBOHYD 627 627
 FT CARBOHYD 649 649
 FT CARBOHYD 1091 1091
 FT CARBOHYD 2038 2038
 FT CARBOHYD 2811 2811
 SQ SEQUENCE 3033 AA: 329165 MW: 79575C1A273BE9E CRC64;
 Query Match 74.7%; Score 712; DB 1; Length 3033;
 Best Local Similarity 69.8%; Pred. No. 5.3e-60;
 Matches 127; Conservative 29; Mismatches 26; Indels 0; Gaps 0;
 QY 1 MAPITAYAQQTGRLGCIITSLTGRDNQVEGVQIVSTAQTFLATCINGVCWTVYHGA 60
 DB 1030 LAPITAYAQQTGRLGIIIVVSMTRGDKTEQAGSIQVLSTVTSFGLTGISGLVTVYHGA 1089
 QY 61 GTRTASPKGPVQMTYNDKDLVGPAPQSGSRSLTPCTCGSSDLYLVTRHADVIPVRR 120
 DB 1090 GNKTLASRGVPTQMTSSAGDVLGVPSPGPTKSLPCTCGAVDLYLVTRHADVIPARR 1149
 QY 121 GDSRGSLLSPRIYLYKSSGGPLLCGAPAGHAVGIFRAAVCTRGVAKAVDFIPVESLEITM 180
 DB 1150 GDKRGALLSPRLSTLKGSSGGPVLCPRGHAVGVFRAAVCAVSRGVAKSIDFIPVETLIDIVT 1209
 QY 181 RS 182
 DB 1210 RS 1211
 RESULT 9
 ID PAAD_PSEAE
 AC Q9HX08; STANDARD; PRT: 209 AA.

DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE 16-OCT-2001 (Rel. 40, Last annotation update)
 GN PA4019.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=287;
 [1]
 SEQUENCE FROM N.A.
 RX STRAIN=ATCC 15692 / PA01;
 MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Miziochi S.D., Warren P.,
 Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 Brody L.L., Coulter S.N., Folger K.K., Kas A., Larbig K., Lim R.M.,
 Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 Reizer J., Salier M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT *Complete genome sequence of Pseudomonas aeruginosa PA01, an
 opportunistic pathogen.;
 RL Nature 406:959-964(2000).
 CC -!- SIMILARITY: BELONGS TO THE POLYPRENYL P-HYDROXYBENZOATE /
 PHENYLACETYLIC ACID DECARBOXYLASES FAMILY.
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE004818; AAG07406.1; -
 DR PIR: H83144; H83144.
 DR InterPro: IPR003182; Flavoprotein.
 DR Pfam: PF02441; Flavoprotein; 1
 KW Hypothetical protein; Lyase; Decarboxylase; Complete proteome.
 SQ SEQUENCE 209 AA: 22367 MW: 01F0081CC495D3F6 CRC64;
 Query Match 9.1%; Score 87; DB 1; Length 209;
 Best Local Similarity 26.2%; Pred. No. 0.28;
 Matches 55; Conservative 18; Mismatches 61; Indels 76; Gaps 12;
 QY 8 AQQTGRLGCIITSLTGRDNQVEGVQ-IVSTAQTFLATCINGVCWTVYHGAQTRTTA 66
 DB 17 AQYGLRLDCLV-----QEEREVHFLISKAAQLVMAT-----ETDVA 53
 QY 67 SPKGP-----VIQMTYNDKDLVGPAPQSGSRSLTP-----CTCGSSDL 105
 DB 54 LPAPQMAQAFLEYCGAAAGQIVFGQND-----HWAPPAGSSAPNAPWICPSTGTL 108
 QY 106 -----YLVTRHADVIPVRRGDSRGLSPR--PIS-----YLYKSSGGPLLCRA 148
 DB 109 SAVATGACNNLIERAADVALKER----RPLVLVPREAPFSSIHLENMLKLSNLGAVILPA 164
 QY 149 GHAVGIFRAAVCTRGVAKAVDFIPVESLET 178
 DB 165 --APGFTHQ----POSVEDLVDFVVARILNT 189
 RESULT 10
 ID HHOA_ARATH
 AC Q9SEL7; Q49507; STANDARD; PRT: 321 AA.
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Protease RhoA, chloroplast precursor (EC 3.4.21.-).
 GN HHOA OR AT4G18370 OR F28J12.30
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

CC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lensch M.H.A., Sokolenko A., Hertmann R.G.;
 RT "Identification and characterization of the chloroplast H₂O₂ protease.
 RT a homolog to the bacterial periplasmic protease H₂O₂.";
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=20083488; PubMed=10617198;
 RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
 RA Pohl T., Duesterhoeft A., Stickema W., Entian K.-D., Terryn N.,
 RA Harris B., Ansonge W., Brandt P., Grivell L., Rieger M.,
 RA Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
 RA Kreis M., Deleney M., Puigdomenech P., Watson M., Schmidheini T.,
 RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
 RA Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,
 RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
 RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
 RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
 RA Weitzenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
 RA Holzner E., Brandt A., Peters S., van Staveren M., Dirks W.,
 RA Mooljman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
 RA Berner S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
 RA De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,
 RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
 RA Clark L., Dognett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,
 RA Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
 RA Borkova D., Bloecker H., Scharfe M., Grimm M., Loehner T.-H.,
 RA Dose S., de Haan M., Maarse A., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Farnmann B., Granderath K., Dauner D., Herzl A.,
 RA Neumann S., Argiridou A., Vitale D., Liguori R., Piravandi E.,
 RA Massenet O., Quigley F., Clabaud G., Muendlein A., Feilber R.,
 RA Schnabl S., Hiller R., Schmidt W., Lechary A., Aubourg S.,
 RA Chedor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
 RA Gibbons T., Weber N., Vandenbol M., Barges M., Totol J., Torres A.,
 RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacor D., Jesse T.,
 RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Biele C.,
 RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
 RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
 RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
 RA Sekhon M., Murray J., Sheel P., Cordes M., Abu-Threideh J.,
 RA Stoneking T., Kallio J., Graves T., Harmon G., Edwards J.,
 RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
 RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
 RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
 RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
 RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
 RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
 RA Ma P., Zhong J., Preston R., Vili D., Shekher M., Matero A., Shah R.,
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
 RA Grant S., Shohdy N., Hasegawa A., Hancock A., Lodhi M., Johnson A.,
 RA Chen E., Marra M., Martinsson R., McCombie W.R.;
 RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
 RT thaliana";
 RL Nature 402:769-777(1999).
 RN [3]
 RP SEQUENCE OF 72-82; 96-110; 150-159; 178-211 AND 306-320.
 RA Schubert M., Peterson U., Funk C., Haas B., Schroeder W.P.,
 RA Kieselbach T.;
 RT "The chloroplast lumen from Arabidopsis thaliana";
 RL Submitted (JUL-2001) to the SWISS-PROT data bank.
 CC -1- SUBCELLULAR LOCATION: Chloroplast; within the thylakoid lumen.
 CC -1- SIMILARITY: BELONGS TO PP2C/FAM112 FAMILY S2C.
 CC -1- CAUTION: Ref.2 sequences differ from that shown due to erroneous
 CC gene model prediction. At4G18370 and At4G18375 were originally
 CC fused into a single gene.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 DR EMBL; AF114386; AAF24060.1;
 DR EMBL; AL021710; CAA16717.1; ALI_SEQ.
 DR EMBL; AL161548; CAB78839.1; ALI_SEQ.
 DR MEROPS; S01.279;
 DR InterPro; IPR001940; Protease2C.
 DR InterPro; IPR001254; Ser_protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PRO0834; PROTEASES2C.
 KW Hydrolyase; Serine protease; Chloroplast; Thylakoid; Transit peptide.
 FT TRANSIT 1 26
 FT TRANSIT 27 71
 FT TRANSIT 72 321
 FT DOMAIN 77 87
 FT ACT_SITE 145 145
 FT ACT_SITE 186 186
 FT ACT_SITE 264 264
 FT ACT_SITE 40 40
 FT CONFLICT 40 40 R -> G (IN REF. 1).
 SQ SEQUENCE 321 AA; 34691 MW; 68DB81E0BD27A7A7 CRC64;
 Query Match 8.8%; Score 84; DB 1; Length 321;
 Best Local Similarity 22.2%; Pred. No. 0.88;
 Matches 48; Conservative 26; Mismatches 60; Indels 82; Gaps 11;
 QY 22 LTGRDKNOVEGEVQIVSTAAQTFLATCINGVCW-----TVYH----- 58
 DB 117 LTDEENGKIEGTG-----SGFVWDLKGLHVTNHYVIAKLTQDGLQCK 161
 QY 59 -----GAGRTIASPKGVQIVQVNTVDKLVGMPAPQGSRSITPCTGSSDLYLVTRHAD 113
 DB 162 VSLVDAGTR--FSKEGKIVGL--DPDNDLAVLKIEGRELNPVVLGTSNDRVQSCF 217
 QY 114 VIPRRRGDSRG-----SILSPRPISYLK-----GSSGGPLCPA 148
 DB 218 AI-----GNPYGYENTLTGVVSGLGREIPSPNGKSISEAIOTDADINSNGSGPLDLSY 272
 QY 149 GHAVGFRAVAVCTR---GVAKAYDF-IPVESLETTM 180
 DB 273 GHIGV-NTATFTRKSGMSSGVNFAIPDVTVRTV 307
 RESULT 11
 ID AAMP_HUMAN STANDARD; PRT; 452 AA.
 AC Q13685;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DE 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Angio-associated migratory cell protein.
 CN AAMP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Brain;
 RX MEDLINE=95262124; PubMed=7743515;
 RA Beckner M.E., Kruttsch H.C., Stracke M.L., Williams S.T.,
 RA Gallardo J.A., Liotta L.A.;
 RT "Identification of a new immunoglobulin superfamily protein expressed
 RT in blood vessels with a heparin-binding consensus sequence.";
 RL Cancer Res. 55:2140-2149(1995).
 CC -1- FUNCTION: MAY HAVE A FUNCTION IN MIGRATING CELLS.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN BLOOD VESSELS. STRONGLY EXPRESSED
 CC IN ENDOTHELIAL CELLS, CYTOTROPHOBLASTS, AND POORLY DIFFERENTIATED
 CC COLON ADENOCARCINOMA CELLS FOUND IN LYMPHATICS.
 CC -1- SIMILARITY: Contains 8 WD repeats.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; M95627; AAA68889.1; -
 DR PIR; I39383; I39383.
 DR Genew; HGNC:18; AAMP.
 DR MIN; 603488; -
 DR GO; GO:0008201; F:heparin binding activity; TAS.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00400; WD40; 8.
 DR SMART; SM00320; WD40; 8.
 DR PROSITE; PS00678; WD_REPEATS_1; 1.
 DR PROSITE; PS50082; WD_REPEATS_2; 6.
 DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
 KW Repeat; WD repeat.
 FT DOMAIN 14 18 HEPARIN-BINDING (POTENTIAL).
 FT DOMAIN 71 77 POLY-GLU.
 FT REPEAT 107 138
 FT REPEAT 150 180
 FT REPEAT 190 220
 FT REPEAT 231 261
 FT REPEAT 276 306
 FT REPEAT 333 363
 FT REPEAT 374 404
 FT REPEAT 416 446
 SQ SEQUENCE 452 AA; 49015 MW; DA1413D25EB236C0 CRC64;

Query Match 8.6%; Score 82; DB 1; Length 452;
 Best Local Similarity 25.3%; Pred. No. 2;
 Matches 42; Conservative 13; Mismatches 47; Indels 64; Gaps 9;
 QY 54 WTIVYAGTTRTIAIPKGPVQIMTVNDKLVGPAPOGSRSL-----TPCTGSSDLYLV 108
 DB 197 WMEWH-----PRAPVLLAGT-ADGNTWMKVPNGDKCTTQGGNCPATCG----- 240
 QY 109 TRHADVIVPVR-----GDSRGS-----LLSPRPISYLGSSG--GPILCPA----- 148
 DB 241 -----VLPDGRKRAVVGVEDGTIRIWLKQSPHVLKGTGEGHQLTCVAAANGDGLILT 295
 QY 149 -----GHAVGIFR-----AAVCTRGVAKAVDFIPVESL 176
 DB 296 GSVDCQAKLVSAITTKVGVGVFRPETVASQPSGLGEESESNSVESL 341

RESULT 12
 Y136_TREPA
 ID Y136_TREPA STANDARD; PRT; 485 AA.
 AC O83172;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical lipoprotein TP0136 precursor.
 GN TP0136.
 OS Treponema pallidum.
 OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
 OX NCBI_TaxID=160;
 RN [1]
 RC STRAIN=Nichols;
 RX MEDLINE=98332770; PubMed=9665876;
 RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
 RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
 RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
 RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
 RA McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,
 RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
 RA Venter J.C.;
 RC "Complete genome sequence of Treponema pallidum, the syphilis

RT spirochete.";
 RL Science 281:375-388(1998).
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
 CC (Potential).
 CC -!- SIMILARITY: BELONGS TO THE TP013X FAMILY OF LIPOPROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; AE001199; AAC65137.1; ALT_INIT.
 DR TIGR; TP0136; -
 KW Hypothetical protein; Lipoprotein; Membrane; Signal;
 KW Complete proteome.
 FT CHAIN 1 23 POTENTIAL.
 FT LIPID 24 485 N-ACYL DIGLYCERIDE (POTENTIAL).
 FT DOMAIN 164 178 GLY/SER-RICH.
 FT DOMAIN 196 210 GLY/SER-RICH.
 FT DOMAIN 253 267 GLY/SER-RICH.
 FT DOMAIN 318 327 POLY-SER.
 FT DOMAIN 444 447 POLY-SER.
 SQ SEQUENCE 485 AA; 48984 MW; C7A4CEDC7DC5CED CRC64;

Query Match 8.6%; Score 82; DB 1; Length 485;
 Best Local Similarity 24.2%; Pred. No. 2;
 Matches 44; Conservative 13; Mismatches 65; Indels 60; Gaps 8;
 QY 23 TGRDKNOVEGEVQIVSTAQTFLATCI--NGVCWTVYHGAG---TRTIAIPKGPVQIMYT 77
 DB 86 TDSK-----KMSIATDGTNTFVLACVPGTGVYKHCVNGAGSSSTGTASPSTETCSQHA 140
 QY 78 NVYDKLNG-----WPAQGSRSLSLTPTCT-----GSSDLYLVTRHADVIP-----VR 118
 DB 141 T-----LVGGTSKPLVLPVGGTNGNCGCGGGGGSSSSSCIIHLVLPVGGTNGNCG 196
 QY 119 RRGDSRGLSLSPRISYK-----GSSGGPLLCAPAGHA 151
 DB 197 CGGGGGSSSSSSSCIIHKVENTDQFLDMGEGYVTTKHLTKNGSSSAGPAQCPCGGG 256
 QY 152 VG 153
 DB 257 GG 258

RESULT 13
 DEGL_ARATH
 ID DEGL_ARATH STANDARD; PRT; 437 AA.
 AC O22609; Q9LK95;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Protease Do-like 1, chloroplast precursor (EC 3.4.21.-).
 GN DEGP1 OR DEGP OR AT3G27925 OR K16N12.18.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 OX NCBI_TaxID=3702;
 RN [1]
 RC SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RX MEDLINE=98175982; PubMed=9507020;
 RA Itzhaki H., Naveh L., Lindahl M., Cook M., Adam Z.;
 RT "Identification and characterization of Degp, a serine protease
 RT associated with the luminal side of the thylakoid membrane.";
 RL J. Biol. Chem. 273:7094-7098(1998).
 RN [2]
 RC SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;

Db 210 PK--NKLRPYPGVGSADLLVGQKVFAGNFGDLHTLTGTGVISLRREIS--SAATGRPI 265

QY 134 SYL-----KGSSGGPLICPAGHAVGIIFRAAVCTRCVAKAVDF-IPVESL 176
:
:|||||||:
:|||||||:

Dd 266 QDVITQDAAINFGSGGPLDSSGTLGINTALTYPSGASSGVGFSPVDIV 317
:
:|||||||:
:|||||||:

RESULT 14

CAC3_BOVIN ID CAC3_BOVIN STANDARD; PRT; 253 AA.

AC P05805;

DT 01-NOV-1988 (Rel. 09, Created)

DI 15-DEC-1998 (Rel. 37, Last sequence update)

DE 13-SEP-2003 (Rel. 42, Last annotation update)

DE Proproteinase E precursor [procarboxypeptidase A complex component III] (Procarboxypeptidase A-S6 subunit III) (PROCPA-S6 III).

DE Bos taurus (Bovine). OS

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea; Bovidae; Bovinae; Bos.

OC NCBI_TaxID=9913;

OX [1]

RX SEQUENCE OF 1-25.

RX MEDLINE=91099520; PubMed=2269366;

RA Pascual R., Vendrell J., Aviles F.X., Bonicel J., Wicker C., Puigserver A.;

RA "Autolysis of proproteinase E in bovine procarboxypeptidase A ternary complex gives rise to subunit III.";

RT FEBS Lett. 277:37-41(1990).

RN [2]

RN SEQUENCE OF 14-253, AND DISULFIDE BONDS.

RX MEDLINE=86220198; PubMed=3519215;

RA Venot N., Sciaky M., Puigserver A., Desnuelle P., Laurent G.;

RT *Amino acid sequence and disulfide bridges of subunit III, a defective endopeptidase present in the bovine pancreatic 6 S procarboxypeptidase A complex.";

RL Eur. J. Biochem. 157:91-99(1986).

RN [3]

RX X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).

RX MEDLINE=9422022; PubMed=8168476;

RA Pignol D., Gaboriau C., Michon T., Kerfelec B., Chapus C., Fontecilla-Camps J.C.;

RT "Crystal structure of bovine procarboxypeptidase A-S6 subunit III, a highly structured truncated zymogen E.";

RI EMBO J. 13:1763-1771(1994).

CC -1- FUNCTION: DEFECTIVE ELASTASE-LIKE SERINE PROTEASE. DOES NOT SEEM TO HAVE A PROTEASE ACTIVITY. ITS LIKELY FUNCTION IS TO PROTECT PROCARBOXYPEPTIDASE A AGAINST DENATURATION IN THE ACIDIC ENVIRONMENT OF THE RUMINANT DUODENUM.

CC -1- SUBUNIT: HETEROTRIMER OF SUBUNIT III; CARBOXYPEPTIDASE A AND CHYMOTRYPSINOGEN C.

CC -1- SUBCELLULAR LOCATION: Extracellular.

CC -1- TISSUE SPECIFICITY: Pancreas.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.

DR PDB; LFON; 14-OCT-96.

DR PDB; LPYT; 27-JAN-97.

DR MEROPS; S01.983; .

DR InterPro; IPRO01314; Chymotrypsin.

DR InterPro; IPRO01254; Ser-protease_Try.

DR Pfam; PF00089; trypsin; 1.

DR PRINTS; PR00722; CHYMOTRYPSIN.

DR SMART; SM00020; TRY_PPC; 1.

DR PROSITE; PS00240; TRYPsin_DOM; 1.

DR PROSITE; PS00134; TRYPsin_HIS; 1.

DR PROSITE; PS00135; TRYPsin_SER; 1.

KW Serine protease homolog; Pancreas; Digestion; 3D-structure.

FT PROPEP 1 11 ACTIVATION PEPTIDE.

FT CHAIN 12 253 PROTEINASE E.

FT DISULFID 41 57

FT DISULFID 100 103

FT DISULFID 140 206

FT DISULFID 171 187

FT DISULFID 196 227

```
FT STRAND 23 33
FT TURN 34 35
FT STRAND 36 41
FT STRAND 44 47
FT TURN 48 49
FT STRAND 50 53
FT HELIX 55 57
FT TURN 60 61
FT STRAND 64 72
FT TURN 73 74
FT STRAND 75 84
FT TURN 87 88
FT STRAND 90 92
FT TURN 94 95
FT TURN 98 99
FT HELIX 101 103
FT STRAND 108 111
FT TURN 119 120
FT STRAND 126 126
FT TURN 130 131
FT TURN 136 137
FT STRAND 139 144
FT TURN 146 147
FT STRAND 157 157
FT STRAND 159 166
FT HELIX 168 171
FT TURN 172 172
FT TURN 174 177
FT HELIX 178 180
FT TURN 183 184
FT STRAND 185 188
FT TURN 196 197
FT TURN 200 201
FT STRAND 203 207
FT TURN 209 210
FT STRAND 213 221
FT STRAND 223 223
FT TURN 224 225
FT STRAND 226 226
FT TURN 230 231
FT STRAND 234 238
FT HELIX 239 241
FT HELIX 243 253
SQ SEQUENCE 253 AA; 27337 MW; 24663724D8AE409C CRC64;

Query Match 7.98; Score 75; DB 1; Length 253;
Best Local Similarity 22.78; Pred. No. 4.9;
Matches 46; Conservative 23; Mismatches 72; Indels 62; Gaps 10;

QY 17 CIITSLTGR-----DKNOVEGEVOIVS-TAAQTFLATCINGVCWTIVYHCAGTRTIA---66
DB 57 CISTRTYQVVLGYEYDSVLEGESEQVPIINGDLFVHPLWNSVCVACGNDIALVLRSA 116
QY 67 -----SPKGPVI-----QMTYTNVDKDLVGPAPQG-SRSLTPCT---99
DB 117 QLGDVKVOLANLPPAGDILPNEAPCYISCGWGLYT-----GGPLPKQLQALLPVVDYE 169
QY 100 -CGSDLYLVTRHADVPVRRGDSRGLSPRPISYLVKSGSGPLCPAG-----HAV 152
DB 170 HCSQMDWMTVKKIM--VCAGGDTR-----SGCNGDSGCPNCPAAGDSWQVHGV 218
QY 153 GIFRAAVCTRGVAKAVDFIPVES 175
DB 219 TSFVSAGCNTIKKPTVTRVSA 241
```

RESULT 15

```
2P3_RABIT
ID 2P3_RABIT STANDARD; PRT: 415 AA.
AC P48R33;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
```

```
DE Zona pellucida sperm-binding protein 3 precursor (Zona pellucida
DE glycoprotein ZP3) (Sperm receptor) (Zona pellucida protein C)
DE (Fragment).
GN ZP3 OR ZPC.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX MEDLINE=95143578; PubMed=7841460;
RA Harris J.D., Hibler D.W., Fontenot G.K., Hsu K.T., Yurewicz E.C.,
RA Sacco A.G.;
RT "Cloning and characterization of zona pellucida genes and cDNAs from
RT a variety of mammalian species: the ZPA, ZPB and ZPC gene families.";
RL DNA Seq. 4:361-393(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U05782; AAA74392.1; -.
DR PIR: S70401; S70401.
DR InterPro: IPR001507; Endoglin/CD105.
DR Pfam: PF00100; zona_pellucida; 1.
DR SMART: SM00241; ZP; 1.
DR PROSITE: PS00682; ZP_DOMAIN; 1.
KW Glycoprotein; Signal; Sulfation; Sperm; Receptor; Transmembrane;
KW Extracellular matrix; Multigene family.
FT NON_TER 1 18
FT SIGNAL <1 18
FT CHAIN 19 415
FT DOMAIN 19 378
FT TRANSMEM 379 399
FT DOMAIN 400 415
FT DOMAIN 41 301
FT ZP.
SQ SEQUENCE 415 AA; 44987 MW; 77396CF1BA3F5CB CRC64;
```

Query Match 7.88; Score 74.5; DB 1; Length 415;
Best Local Similarity 26.78; Pred. No. 9.5;
Matches 31; Conservative 17; Mismatches 53; Indels 15; Gaps 5;

```
QY 55 TVYHCAGTRTITASPKGP-VIQMYTNVDKDLVGPAPQGSRLTPCTCGSSDLYLVTRHAD 113
DB 271 TVYITCHLVTPAQQAQADRLNKACSFNQSSSWAPVEGSADICEC-CGNGDCDLIAGS-- 327
QY 114 VIPVRRGDSRGLSPRPISYLVKSGSGPL--LCPAGHAVGIFRAAVCTRGVAKA 167
DB 328 --PKNQNHAAARSLRRRHVTEADVTVGPLIFLGKAGDPAG-----TEGLASA 374
```

Search completed: August 30, 2003, 19:13:43
Job time : 13.0138 secs

QY 122 DSRGSLSPRPISYLKSGGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVLESLETTMR 181
 DB 121 DSRGSLSPRPISYLKSGGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTMR 180

QY 182 S 182
 DB 181 S 181

RESULT 2
 Q91RTS
 ID Q91RTS PRELIMINARY; PRT: 181 AA.
 AC Q91RTS
 DT 01-DEC-2001 (TRENBLrel. 19, Created)
 DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE NS3 protease (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Pt.4;
 RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
 RT "Genetic Diversity and response to IPI of the NS3 Protease Gene from
 RT Clinical Strains of the Hepatitis C Virus."
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF369218; AAK3453.1; -
 DR InterPro: IPR004109; HCV_NS3.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Protease.
 KW Protease.
 FT NON_TER 1
 FT NON_TER 181
 SQ SEQUENCE 181 AA; 19130 MW; 85D91869299B7C35 CRC64;

Query Match 99.2%; Score 945; DR 12; Length 181;
 Best Local Similarity 99.4%; Pred. No. 1.9e-88;
 Matches 180; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 APITAYAOOTRGLGCIITSLTGRDKNQVEGEVOIVSTAQTFLATCINGVCTVYHGAG 61
 DB 1 APITAYAOOTRGLGCIITSLTGRDKNQVEGEVOIVSTAQTFLATCINGVCTVYHGAG 60

QY 62 TRTIASPKGPVQIYNTVDKLVGWPAQPSRSITPCTCGSSDLYLVTRHADVIPVRRRG 121
 DB 61 TRTIASPKGPVQIYNTVDKLVGWPAQPSRSITPCTCGSSDLYLVTRHADVIPVRRRG 120

QY 122 DSRGSLSPRPISYLKSGGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVLESLETTMR 181
 DB 121 DSRGSLSPRPISYLKSGGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTMR 180

QY 182 S 182
 DB 181 S 181

RESULT 3
 Q81756
 ID Q81756 PRELIMINARY; PRT: 2436 AA.
 AC Q81756;
 DT 01-NOV-1996 (TRENBLrel. 01, Created)
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Choo Q.-L., Richman K., Han J.;
 RA Choo Q.-L., Richman K., Han J.;
 RT "The nucleotide sequence of the Hepatitis C viral genome.";

RL Submitted (MAY-1990) to the EMBL/GenBank/DBJ databases.
 DR EMBL: M32084; AAA45677.1; -
 DR HSSP: P27958; 1A1V.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRP.
 DR InterPro: IPR001650; Helicase.C.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; helicase.C; 1.
 DR Pfam: PF00998; Viral_RdRP; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR PROSITE; PS50507; RDRP_POSITIVE; 1.
 DR PROSITE; PS50521; RDRP_VIRAL; 1.
 KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
 KW Hydrolase; Nonstructural protein; Polyprotein;
 KW RNA-directed RNA polymerase; Transferase; Transmembrane.
 FT NON_TER 1
 FT NON_TER 2436
 SQ SEQUENCE 2436 AA; 264734 MW; D7B9872900BE3125 CRC64;

Query Match 99.0%; Score 943; DB 12; Length 2436;
 Best Local Similarity 98.4%; Pred. No. 7.1e-87;
 Matches 179; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPITAYAOOTRGLGCIITSLTGRDKNQVEGEVOIVSTAQTFLATCINGVCTVYHGA 60
 DB 576 LAPITAYAOOTRGLGCIITSLTGRDKNQVEGEVOIVSTAQTFLATCINGVCTVYHGA 635

QY 61 GTRTASPKGPVQIYNTVDKLVGWPAQPSRSITPCTCGSSDLYLVTRHADVIPVRRR 120
 DB 636 GTRTASPKGPVQIYNTVDKLVGWPAQPSRSITPCTCGSSDLYLVTRHADVIPVRRR 695

QY 121 GDSRGSLLSPRPISYLKSGGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVLESLETTM 180
 DB 696 GDSRGSLLSPRPISYLKSGGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 755

QY 181 RS 182
 DB 756 RS 757

RESULT 4
 Q91FE5
 ID Q91FE5 PRELIMINARY; PRT: 3011 AA.
 AC Q91FE5;
 DT 01-OCT-2000 (TRENBLrel. 15, Created)
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE Genome polyprotein.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-21262212; PubMed-11369872;
 RA Lanford R.E., Lee H., Chavez D., Guerra B., Brasky K.M.;
 RT "Infectious cDNA clone of the hepatitis C virus genotype 1 prototype
 RT sequence."
 RL J. Gen. Virol. 82:1291-1297(2001).

```
CC CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: AF2711632; AAF81759.1; -.
DR HSSP: P27958; 1A1V.
DR InterPro: IPR000345; CytC_heme_bind.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_err.
DR InterPro: IPR002531; HCV_NSI.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RDRP.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_err; 1.
DR Pfam: PF01560; HCV_NSI; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RDRP; 1.
DR ProDom: PD186062; HCV_NSI; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS00190; CYTOCHROME_C; 1.
DR PROSITE: PS0507; RDRP_POSITIVE; 1.
DR PROSITE: PS0521; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA; 327124 MW; 2489CE74AC864E58 CRC64;

Query Match 99.0%; Score 943; DB 12; Length 3011;
Best Local Similarity 98.4%; Pred. No. 9.2e-87;
Matches 179; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPITAYAAQQTGRLGCIITSITGRDKNOVEGEVOIVSTAAQTFLATCINGVCWTVYHGA 60
:|||||
Db 1026 LAPITAYAAQQTGRLGCIITSITGRDKNOVEGEVOIVSTAAQTFLATCINGVCWTVYHGA 1085

QY 61 GRTTIASPKGPVIOMYTNVDKDLVGPAPQGSRLTPTCGSSDLYLVTRHADVIPVRRR 120
|||||
Db 1086 GRTTIASPKGPVIOMYTNVDQDLVGPAPQGSRLTPTCGSSDLYLVTRHADVIPVRRR 1145

QY 121 GDSRGSLLSPRPISYLVKSGSGPLLCPCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 180
|||||
Db 1146 GDSRGSLLSPRPISYLVKSGSGPLLCPCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 1205

QY 181 RS 182
||
Db 1206 RS 1207

RESULT 5
Q91RR3 PRELIMINARY; PRT; 181 AA.
ID AC Q91RR3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

Query Match 99.0%; Score 943; DB 12; Length 3011;
Best Local Similarity 98.4%; Pred. No. 9.2e-87;
Matches 179; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPITAYAAQQTGRLGCIITSITGRDKNOVEGEVOIVSTAAQTFLATCINGVCWTVYHGA 60
:|||||
Db 1026 LAPITAYAAQQTGRLGCIITSITGRDKNOVEGEVOIVSTAAQTFLATCINGVCWTVYHGA 1085

QY 61 GRTTIASPKGPVIOMYTNVDKDLVGPAPQGSRLTPTCGSSDLYLVTRHADVIPVRRR 120
|||||
Db 1086 GRTTIASPKGPVIOMYTNVDQDLVGPAPQGSRLTPTCGSSDLYLVTRHADVIPVRRR 1145

QY 121 GDSRGSLLSPRPISYLVKSGSGPLLCPCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 180
|||||
Db 1146 GDSRGSLLSPRPISYLVKSGSGPLLCPCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 1205

QY 181 RS 182
||
Db 1206 RS 1207

RESULT 5
Q91RR3 PRELIMINARY; PRT; 181 AA.
ID AC Q91RR3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
```

```
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pt.4B;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF369240; AAK54565.1; -.
DR InterPro: IPR004109; HCV_NS3.
DR Pfam: PF02907; HCV_NS3; 1.
KW Protease.
KW NON_TER 1 1
FT NON_TER 181 181
SQ SEQUENCE 181 AA; 19115 MW; 5D85F88AD7AC1A11 CRC64;

Query Match 98.8%; Score 942; DB 12; Length 181;
Best Local Similarity 98.9%; Pred. No. 3.8e-88;
Matches 179; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 APITAYAAQQTGRLGCIITSITGRDKNOVEGEVOIVSTAAQTFLATCINGVCWTVYHGA 61
|||||
Db 1 APITAYAAQQTGRLGCIITSITGRDKNOVEGEVOIVSTAAQTFLATCINGVCWTVYHGA 60

QY 62 TRTIASPKGPVIOMYTNVDKDLVGPAPQGSRLTPTCGSSDLYLVTRHADVIPVRRRG 121
|||||
Db 61 TRTIASPKGPVIOMYTNVDRLVGPAPQGSRLTPTCGSSDLYLVTRHADVIPVRRRG 120

QY 122 DSRGSLSPRPISYLVKSGSGPLLCPCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMR 181
|||||
Db 121 DSRGSLSPRPISYLVKSGSGPLLCPCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMR 180

QY 182 S 182
|
Db 181 S 181

RESULT 6
Q91RS1 PRELIMINARY; PRT; 181 AA.
ID AC Q91RS1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pt.K;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF369232; AAK54557.1; -.
DR InterPro: IPR004109; HCV_NS3.
DR Pfam: PF02907; HCV_NS3; 1.
KW Protease.
KW NON_TER 1 1
FT NON_TER 181 181
SQ SEQUENCE 181 AA; 19114 MW; ABB90B5B3ABA4E26 CRC64;

Query Match 98.8%; Score 942; DB 12; Length 181;
Best Local Similarity 98.9%; Pred. No. 3.8e-88;
Matches 179; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 APITAYAAQQTGRLGCIITSITGRDKNOVEGEVOIVSTAAQTFLATCINGVCWTVYHGA 61
|||||
Db 1 APITAYAAQQTGRLGCIITSITGRDKNOVEGEVOIVSTAAQTFLATCINGVCWTVYHGA 60
```



```

QY 62 TRTIASPKGPVIOMYTNVDKLVGWPAPOGSRSLTPCTCGSSDLYLVTRHADVIPVRRRG 121
DB 1 TRTIASPKGPVIOMYTNVDKLVGWPAPOGSRSLTPCTCGSSDLYLVTRHADVIPVRRRG 120
QY 122 DSRGSLSPRISYLVKSGGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMR 181
DB 121 DSRGSLSPRISYLVKSGGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTMR 180
QY 182 S 182
DB 181 S 181

RESULT 7
Q91R08 PRELIMINARY; PRT; 181 AA.
AC Q91R08;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic diversity and response to IFN of the NS3 Protease Gene from
RT Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF369245; AAK54570.1; -.
DR InterPro; IPR004109; HCV_NS3.
DR Pfam; PF02907; HCV_NS3; 1.
KW Protease.
FT NON_TER 1 181
FT NON_TER 181
SQ SEQUENCE 181 AA; 19144 MW; 614ADA8B0F33CCAF CRC64;

Query Match 98.8%; Score 942; DB 12; Length 181;
Best Local Similarity 98.9%; Pred. No. 3.8e-88;
Matches 179; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 APITAYAOOTRGLGCIITSLTGRDKNQVEGEVIVSTAAQTFLATCINGVCWTVYHGAG 61
DB 1 APITAYAOOTRGLGCIITSLTGRDKNQVEGEVIVSTAAQTFLATCINGVCWTVYHGAG 60
QY 62 TRTIASPKGPVIOMYTNVDKLVGWPAPOGSRSLTPCTCGSSDLYLVTRHADVIPVRRRG 121
DB 61 TRTIASPKGPVIOMYTNVDKLVGWPAPOGSRSLTPCTCGSSDLYLVTRHADVIPVRRRG 120
QY 122 DSRGSLSPRISYLVKSGGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMR 181
DB 121 DSRGSLSPRISYLVKSGGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTMR 180
QY 182 S 182
DB 181 S 181

RESULT 8
Q91RT1 PRELIMINARY; PRT; 181 AA.
AC Q91RT1;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic diversity and response to IFN of the NS3 Protease Gene from
RT Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF369245; AAK54570.1; -.
DR InterPro; IPR004109; HCV_NS3.
DR Pfam; PF02907; HCV_NS3; 1.
KW Protease.
FT NON_TER 1 181
FT NON_TER 181
SQ SEQUENCE 181 AA; 19144 MW; 614ADA8B0F33CCAF CRC64;

Query Match 98.8%; Score 942; DB 12; Length 181;
Best Local Similarity 98.9%; Pred. No. 3.8e-88;
Matches 179; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 APITAYAOOTRGLGCIITSLTGRDKNQVEGEVIVSTAAQTFLATCINGVCWTVYHGAG 61
DB 1 APITAYAOOTRGLGCIITSLTGRDKNQVEGEVIVSTAAQTFLATCINGVCWTVYHGAG 60
QY 62 TRTIASPKGPVIOMYTNVDKLVGWPAPOGSRSLTPCTCGSSDLYLVTRHADVIPVRRRG 121
DB 61 TRTIASPKGPVIOMYTNVDKLVGWPAPOGSRSLTPCTCGSSDLYLVTRHADVIPVRRRG 120
QY 122 DSRGSLSPRISYLVKSGGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMR 181
DB 121 DSRGSLSPRISYLVKSGGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTMR 180
QY 182 S 182
DB 181 S 181

RESULT 9
Q91RR6 PRELIMINARY; PRT; 181 AA.
AC Q91RR6;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic diversity and response to IFN of the NS3 Protease Gene from
RT Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF369245; AAK54570.1; -.
DR InterPro; IPR004109; HCV_NS3.
DR Pfam; PF02907; HCV_NS3; 1.
KW Protease.
FT NON_TER 1 181
FT NON_TER 181
SQ SEQUENCE 181 AA; 19101 MW; 614ADA8B0F33CCAF CRC64;

Query Match 98.6%; Score 940; DB 12; Length 181;
Best Local Similarity 98.3%; Pred. No. 6e-88;
Matches 178; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 APITAYAOOTRGLGCIITSLTGRDKNQVEGEVIVSTAAQTFLATCINGVCWTVYHGAG 61
DB 1 APITAYAOOTRGLGCIITSLTGRDKNQVEGEVIVSTAAQTFLATCINGVCWTVYHGAG 60
QY 62 TRTIASPKGPVIOMYTNVDKLVGWPAPOGSRSLTPCTCGSSDLYLVTRHADVIPVRRRG 121
DB 61 TRTIASPKGPVIOMYTNVDKLVGWPAPOGSRSLTPCTCGSSDLYLVTRHADVIPVRRRG 120
QY 122 DSRGSLSPRISYLVKSGGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMR 181
DB 121 DSRGSLSPRISYLVKSGGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTMR 180
QY 182 S 182
DB 181 S 181

```

```

Db      61  TTTIASPQGPVQIOMTYNDQDLVGVPAPOGARSPTCTCGSSDLYLVTRHADVPVRRRG 120
QY      122  DSRGSLSPRPISYLKSGGGPILCPAGHAGVGFRAAVCTRGVAKAVDFIPVESLETTMR 181
Db      121  DSRGSLSPRPISYLKSGGGPILCPAGHAGVGFRAAVCTRGVAKAVDFIPVESLETTMR 180
QY      182  S 182
Db      181  S 181

RESULT 10
Q91RS9
ID      Q91RS9      PRELIMINARY;      PRT:      181 AA.
AC      Q91RS9;
DT      01-DEC-2001 (TReMBLrel. 19, Created)
DT      01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DE      NS3 protease (Fragment).
OS      Hepatitis C virus.
OC      Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC      Hepacivirus.
OX      NCBI_TaxID=11103;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-Pt.174;
RA      Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT      "Genetic diversity and response to IFN of the NS3 protease gene from
RT      clinical strains of the Hepatitis C virus."
RL      Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR      EMBL: AF369224; AAK54549.1; -;
DR      InterPro: IPR004109; HCV_NS3.
KW      Pfam: PF02907; HCV_NS3; 1.
KW      Protease.
FT      NON_TER
FT      NON_TER
SQ      SEQUENCE 181 AA; 19131 MW; 8BD7FC2769DBD635 CRC64;

Query Match      98.6%; Score 940; DB 12; Length 181;
Best Local Similarity 98.6%; Pred. No. 6e-88;
Matches 179; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      2  APITAYAOQTRGLGCIITSLTGRDKNOVEGEVQIVSTAAQTFLATCINGVCWTVYHGAG 61
Db      1  APITAYAOQTRGLGCIITSLTGRDKNOVEGEVQIVSTAAQTFLATCINGVCWTVYHGAG 60
QY      62  TTTIASPQGPVQIOMTYNDKDLVGVPAPOGARSPTCTCGSSDLYLVTRHADVPVRRRG 121
Db      61  TTTIASPQGPVQIOMTYNDKDLVGVPAPOGARSPTCTCGSSDLYLVTRHADVPVRRRG 120
QY      122  DSRGSLSPRPISYLKSGGGPILCPAGHAGVGFRAAVCTRGVAKAVDFIPVESLETTMR 181
Db      121  DSRGSLSPRPISYLKSGGGPILCPAGHAGVGFRAAVCTRGVAKAVDFIPVESLETTMR 180
QY      182  S 182
Db      181  S 181

RESULT 11
Q91RS3
ID      Q91RS3      PRELIMINARY;      PRT:      181 AA.
AC      Q91RS3;
DT      01-DEC-2001 (TReMBLrel. 19, Created)
DT      01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DE      NS3 protease (Fragment).
OS      Hepatitis C virus.
OC      Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC      Hepacivirus.
OX      NCBI_TaxID=11103;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-HC-J1;
RA      Miyakawa Y., Mayumi M.;
RA      Okamoto H., Okada S., Sugiyama Y., Kuraï K., Izuka H., Machida A.,
RT      "Nucleotide sequences of the genomic RNA of hepatitis C virus isolated
RT      from a human carrier: comparison with reported isolates for conserved
RT      and divergent regions."
RL      J. Gen. Virol. 72:2697-2704(1991).
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN-HC-J1;
RX      MEDLINE=93117120; PubMed=1335573;
RA      Okamoto H., Kanai N., Mishihiro S.;
RT      "Full-length nucleotide sequence of a Japanese hepatitis C virus
RT      isolate (HC-J1) with high homology to USA isolates."
RL      Nucleic Acids Res. 20:6410-6410(1992).
RN      [4]

```

```

RC      STRAIN-Pt.24;
RA      Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT      "Genetic diversity and response to IFN of the NS3 protease gene from
RT      clinical strains of the Hepatitis C virus."
RL      Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR      EMBL: AF369230; AAK54555.1; -;
DR      InterPro: IPR004109; HCV_NS3.
KW      Pfam: PF02907; HCV_NS3; 1.
KW      Protease.
FT      NON_TER
FT      NON_TER
SQ      SEQUENCE 181 AA; 19132 MW; 0BB90B5F3AB95250 CRC64;

Query Match      98.5%; Score 939; DB 12; Length 181;
Best Local Similarity 98.3%; Pred. No. 7.6e-88;
Matches 178; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY      2  APITAYAOQTRGLGCIITSLTGRDKNOVEGEVQIVSTAAQTFLATCINGVCWTVYHGAG 61
Db      1  APITAYAOQTRGLGCIITSLTGRDKNOVEGEVQIVSTAAQTFLATCINGVCWTVYHGAG 60
QY      62  TTTIASPQGPVQIOMTYNDKDLVGVPAPOGARSPTCTCGSSDLYLVTRHADVPVRRRG 121
Db      61  TTTIASPQGPVQIOMTYNDKDLVGVPAPOGARSPTCTCGSSDLYLVTRHADVPVRRRG 120
QY      122  DSRGSLSPRPISYLKSGGGPILCPAGHAGVGFRAAVCTRGVAKAVDFIPVESLETTMR 181
Db      121  DSRGSLSPRPISYLKSGGGPILCPAGHAGVGFRAAVCTRGVAKAVDFIPVESLETTMR 180
QY      182  S 182
Db      181  S 181

RESULT 12
Q03463
ID      Q03463      PRELIMINARY;      PRT:      3011 AA.
AC      Q03463;
DT      01-NOV-1996 (TReMBLrel. 01, Created)
DT      01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT      01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE      Genome polyprotein.
OS      Hepatitis C virus.
OC      Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC      Hepacivirus.
OX      NCBI_TaxID=11103;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-HC-J1;
RX      MEDLINE=91013116; PubMed=2170712;
RA      Okamoto H., Okada S., Sugiyama Y., Yotsumoto S., Tanaka T.,
RA      Yoshizawa H.;
RT      "The 5'-terminal sequence of the hepatitis C virus genome."
RL      Jpn. J. Exp. Med. 60:167-177(1990).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN-HC-J1;
RX      MEDLINE=92044440; PubMed=1658196;
RA      Okamoto H., Okada S., Sugiyama Y., Kuraï K., Izuka H., Machida A.,
RA      Miyakawa Y., Mayumi M.;
RT      "Nucleotide sequences of the genomic RNA of hepatitis C virus isolated
RT      from a human carrier: comparison with reported isolates for conserved
RT      and divergent regions."
RL      J. Gen. Virol. 72:2697-2704(1991).
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN-HC-J1;
RX      MEDLINE=93117120; PubMed=1335573;
RA      Okamoto H., Kanai N., Mishihiro S.;
RT      "Full-length nucleotide sequence of a Japanese hepatitis C virus
RT      isolate (HC-J1) with high homology to USA isolates."
RL      Nucleic Acids Res. 20:6410-6410(1992).
RN      [4]

```

```
RP SEQUENCE FROM N.A.
RC STRAIN-HC-JI;
RA Okamoto H.;
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
RN [5]
RN SEQUENCE FROM N.A.
RC STRAIN-HC-JI;
RX MDLINE=941174722; Pubmed=7510436;
RA Mink M., Benichou S., Madaule P., Tiollais P., Prince A.,
RA Inchausti G.;
RT "Characterization and mapping of a B-cell immunogenic domain in
RT hepatitis C virus E2 glycoprotein using a yeast peptide library.";
RL Virology 200; 246-255(1994).
CC -1- SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL: D10749; BAA01582.1; -.
DR HSP; P27958; IHEI.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_NS5b.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01538; HCV_NS1; 1.
DR Pfam: PF01560; HCV_NS2; 1.
DR Pfam: PF01538; HCV_NS3; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RdRP; 1.
DR PRODOM: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS05021; RDRP_POSITIVE; 1.
DR PROSITE: PS05021; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA; 327112 MW; 97E9052C0250463B CRC64;

Query Match 98.5%; Score 939; DB 12; Length 3011;
Best Local Similarity 97.8%; Pred. No. 2.4e-86;
Matches 178; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAPITAYAAQTGRLGCGIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCTVYHGAG 60
DB 1026 LAPITAYAAQTGRLGCGIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCTVYHGAG 1085
QY 61 GTRTIASPKGPVIOYNTVDKLVGMPAPQGSRLTPTCTGSSDLYLVTRHADVIPVRR 120
DB 1086 GTRTIASPKGPVIOYNTVDKLVGMPAPQGSRLTPTCTGSSDLYLVTRHADVIPVRR 1145
QY 121 GDSRGSLLSPRPISYLKSGSGGPGLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 180
DB 1146 GDSRGSLLSPRPISYLKSGSGGPGLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 1205
QY 181 RS 182
DB 1206 RS 1207
```

```
RESULT 13
Q91RT4
ID Q91RT4 PRELIMINARY; PRT; 181 AA.
AC Q91RT4;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN-Pt.176;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF369219; AAK54544.1; -.
DR InterPro: IPR004109; HCV_NS3.
DR Pfam: PF02907; HCV_NS3; 1.
DR Protease.
KW NON_TER 1 1
FT NON_TER 181
SQ SEQUENCE 181 AA; 19059 MW; 1E53C47AE8B7E5C9 CRC64;

Query Match 98.4%; Score 938; DB 12; Length 181;
Best Local Similarity 98.3%; Pred. No. 9.6e-88;
Matches 178; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 APITAYAAQTGRLGCGIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCTVYHGAG 61
DB 1 APITAYAAQTGRLGCGIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCTVYHGAG 60
QY 62 TRTIASPKGPVIOYNTVDKLVGMPAPQGSRLTPTCTGSSDLYLVTRHADVIPVRR 121
DB 61 TRTIASPKGPVIOYNTVDKLVGMPAPQGSRLTPTCTGSSDLYLVTRHADVIPVRR 120
QY 122 DSRGSLLSPRPISYLKSGSGGPGLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 181
DB 121 DSRGSLLSPRPISYLKSGSGGPGLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 180
QY 182 S 182
DB 181 S 181

RESULT 14
Q91RS8
ID Q91RS8 PRELIMINARY; PRT; 181 AA.
AC Q91RS8;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN-Pt.176;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF369225; AAK54550.1; -.
DR InterPro: IPR004109; HCV_NS3.
DR Pfam: PF02907; HCV_NS3; 1.
KW Protease.
FT NON_TER 1 1
```

Job time : 36.7298 secs

```

FT NON_TER 181 181
SQ SEQUENCE 181 AA: 19114 MW: 574AC47A8AE5D2 CRC64;

Query Match      98.4%; Score 936; DB 12; Length 181;
Best Local Similarity 98.3%; Pred. No. 9.6e-88;
Matches 178; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 APITAYAOOTRGLGCIITSITGRDKNOVEGEVOIVSTAAOTFLATCINGVCWTVYHGAG 61
   |
Db 1 APITAYAOOTRGLGCIITSITGRDKNOVEGEVOIVSTAAOTFLATCINGVCWTVYHGAG 60
   |
QY 62 TRTIASPKGPVIQMTYNVDKDLVGNPAPOGSSITPCTCGSSDLYLVTRHADVIPIVRRRG 121
   |
Db 61 TRTIASPKGPVIQMTYNVDKDLVGNPAPOGSSITPCTCGSSDLYLVTRHADVIPIVRRRG 120
   |
QY 122 DSRGSLSPRPISYLYKSGSGGPLLCPCAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMR 181
   |
Db 121 DSRGSLSPRPISYLYKSGSGGPLLCPCAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTMR 180
   |
QY 182 S 182
   |
Db 181 S 181

```

RESULT 15

```

Q91RT3
ID Q91RT3 PRELIMINARY; PRT; 181 AA.
AC Q91RT3;
DT 01-DEC-2001 (TREMELrel. 19, Created)
DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE NS3 protease (fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_taxid=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=pt.11;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Meyers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF369220; AAK5454.1; -
DR InterPro; IPR004109; HCV_NS3.
DR Pfam; PF02907; HCV_NS3; 1.
KW Protease.
FT NON_TER 1 1
FT NON_TER 181 181
SQ SEQUENCE 181 AA: 19116 MW: 9648807F49EB1D43 CRC64;

```

```

Query Match      98.4%; Score 938; DB 12; Length 181;
Best Local Similarity 98.3%; Pred. No. 9.6e-88;
Matches 178; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 APITAYAOOTRGLGCIITSITGRDKNOVEGEVOIVSTAAOTFLATCINGVCWTVYHGAG 61
   |
Db 1 APITAYAOOTRGLGCIITSITGRDKNOVEGEVOIVSTAAOTFLATCINGVCWTVYHGAG 60
   |
QY 62 TRTIASPKGPVIQMTYNVDKDLVGNPAPOGSSITPCTCGSSDLYLVTRHADVIPIVRRRG 121
   |
Db 61 TRTIASPKGPVIQMTYNVDKDLVGNPAPOGSSITPCTCGSSDLYLVTRHADVIPIVRRRG 120
   |
QY 122 DSRGSLSPRPISYLYKSGSGGPLLCPCAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMR 181
   |
Db 121 DSRGSLSPRPISYLYKSGSGGPLLCPCAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTMR 180
   |
QY 182 S 182
   |
Db 181 S 181

```

Search completed: August 30, 2003, 19:18:17

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 19:02:22 ; Search time 14.9789 Seconds
(without alignments)
1168.492 Million cell updates/sec

Title: US-09-965-594-1

Perfect score: 953

Sequence: 1 MAPITAYAOOTRGLLCIIIT.....GVAKAVDFPVESLFTTMS 162

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_76:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	943	99.0	3011	1	GNWVC3
2	939	98.5	3011	1	genome polyprotein
3	927	97.3	3011	1	genome polyprotein
4	892	93.6	3010	1	genome polyprotein
5	891	93.5	3010	1	genome polyprotein
6	887	93.1	3010	1	genome polyprotein
7	884	92.8	3010	1	genome polyprotein
8	864	90.7	3010	1	genome polyprotein
9	804	84.4	3044	1	genome polyprotein
10	714	74.9	3033	1	genome polyprotein
11	712	74.7	3033	1	genome polyprotein
12	267.5	28.1	3005	2	genome polyprotein
13	255.5	26.8	2970	2	polyprotein - marm
14	87	9.1	209	2	probable aromatic
15	82	8.6	452	2	angio-associated m
16	82	8.6	495	2	hypothetical prote
17	81.5	8.6	476	2	heat shock transcr
18	78.5	8.2	620	2	cytochrome-c oxida
19	78.5	8.2	981	2	beta transducin ho
20	78	8.2	239	2	serine proteinase
21	77.5	8.1	398	2	probable periplasm
22	77	8.1	904	2	hypothetical prote
23	76.5	8.0	270	2	hypothetical prote
24	76.5	8.0	868	2	aconitate hydratase
25	75	7.9	240	1	procarboxypeptidas
26	74.5	7.8	415	2	zona pellucida gly
27	74.5	7.8	755	2	hypothetical prote
28	74.5	7.8	868	2	aconitate hydratase
29	74.5	7.8	911	2	transferrin-bindin

30	74	7.8	140	2	C72705	hypothetical prote
31	74	7.8	377	2	A75335	hypothetical prote
32	74	7.8	808	2	G86208	protein F22G5.28 I
33	73.5	7.7	356	2	F90978	hypothetical prote
34	73.5	7.7	447	2	S76033	hypothetical prote
35	73.5	7.7	451	2	H82044	C4-dicarboxylate t
36	73.5	7.7	566	2	H84203	phosphate ABC tran
37	73.5	7.7	846	2	T04533	hypothetical prote
38	73.5	7.7	910	2	C81832	transferrin-bindin
39	73.5	7.7	915	2	F81196	transferrin-bindin
40	73.5	7.7	1820	2	A55494	latent transferrin
41	73	7.7	239	2	A89967	serine proteinase
42	73	7.7	354	2	T49806	hypothetical prote
43	73	7.7	433	2	H97199	htrA-like serine p
44	73	7.7	535	2	S65762	chitinase (EC 3.2.
45	72.5	7.6	249	2	A55634	granzyme M (EC 3.4

ALIGNMENTS

RESULT 1

GNWVC3

genome polyprotein - hepatitis C virus (strain HCV-1)

N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain HCV-1) (nonst

protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

C>Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 19-Jan-2001

C:Accession: A39166; P00403; P00404

R:Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.

Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991

A:Title: Genetic organization and diversity of the hepatitis C virus.

A:Reference number: A39166; MUID:91172826; PMID:1848704

A:Accession: A39166

A:Molecule type: mRNA

A:Residues: 1-3011 <CHOS>

A:Cross-references: GB:M62321; NID:9329873; PIDN:AAA45676.1; PID:9329874

R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap

J. Gen. Virol. 73, 1131-1141, 1992

A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship

A:Reference number: P00393; MUID:92268871; PMID:1316939

A:Accession: P00403

A:Molecule type: genomic RNA

A:Residues: 1577-1633 <CHAS>

A:Cross-references: DDBJ:D10128

A:Experimental source: Isolates E-b16

A:Accession: P00404

A>Status: preliminary

A:Molecule type: genomic RNA

A:Residues: 1577-1633 <CH2>

A:Experimental source: Isolates E-b17

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstru

F:1-115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: envelope protein M #status predicted <EPM>

F:192-389/Product: major envelope protein E #status predicted <MEE>

F:390-129/Product: nonstructural protein NS1 #status predicted <NS1>

F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F:1007-1515/Product: hepatitis C virus genome polyprotein

F:1230-1237/Product: nucleotide-binding motif A (P-loop)

F:1312-1317/Region: nucleotide-binding motif A (P-loop)

F:1316-1319/Region: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>

F:1863-2013/Product: nonstructural protein NS5 #status predicted <NS5>

F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,20

Query Match 99.0% Score 943; DB 1; Length 3011;

Best Local Similarity 98.4% Pred. No. 5.7e-81;

Matches 179; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Oy 1 MAPITAYAOOTRGLLCIIITSLTGRKNOVEGOIVSTAQTFLATCINGVCWTVYHGA 60

|||||

Db 1026 LAPITAYAAQOTRGLGCIITSLTRDKNQVEGEVQIVSTAATQIFLATCINGVCWTYVHGA 1085

QY 61 GTRTIASPKGPVQMYTNVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 120
 |||||
 Db 1086 GTRTIASPKGPVQMYTNVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 1145
 |||||

QY 121 GDSRGSLLSPRPISYLKSGSGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 180
 |||||

Db 1146 GDSRGSLLSPRPISYLKSGSGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 1205
 |||||

QY 181 RS 182
 ||

Db 1206 RS 1207

RESULT 2

S40770
 genome polyprotein - hepatitis C virus
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain H) (nonstruc)
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
 C:Accession: S40770; PC1285
 R:Okamoto, H.: Submitted to the EMBL Data Library, March 1992
 A:Reference number: S40770
 A:Accession: S40770
 A:Molecule type: genomic RNA
 A:Residues: 1-3011 <OK>
 A:Cross-references: EMBL:D10749; NID:q221586; PIDN:BA001582.1; PID:q221587
 R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.: Tsuda,
 Jpn. J. Exp. Med. 60, 167-177, 1990
 A:Title: The 5'-terminal sequence of the hepatitis C virus genome.
 A:Reference number: PC1284; MUID:91013116; PMID:2170712
 A:Accession: PC1285
 A:Molecule type: genomic RNA
 A:Residues: 1-513 <OK>
 A:Cross-references: GB:D00831; NID:q221511; PIDN:BA00705.1; PID:q221512
 A:Experimental source: isolate HC-J1
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
 F:2-115/Product: capsid protein C #status predicted <GPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <WEE>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: hepatitis C virus genome
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B
 F:1316-1319/Region: DEXH motif

Query Match 98.5%; Score 939; DB 1; Length 3011;
 Best Local Similarity 97.8%; Pred. No. 1.4e-80;
 Matches 178; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAPITAYAAQOTRGLGCIITSLTRDKNQVEGEVQIVSTAATQIFLATCINGVCWTYVHGA 60
 :|||||

Db 1026 LAPITAYAAQOTRGLGCIITSLTRDKNQVEGEVQIVSTAATQIFLATCINGVCWTYVHGA 1085
 :|||||

QY 61 GTRTIASPKGPVQMYTNVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 120
 |||||

Db 1086 GTRTIASPKGPVQMYTNVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 1145
 |||||

QY 121 GDSRGSLLSPRPISYLKSGSGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 180
 |||||

Db 1146 GDSRGSLLSPRPISYLKSGSGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 1205
 |||||

QY 181 RS 182
 ||

Db 1206 RS 1207

Query Match 97.3%; Score 927; DB 1; Length 3011;
 Best Local Similarity 96.2%; Pred. No. 1.9e-79;
 Matches 175; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 MAPITAYAAQOTRGLGCIITSLTRDKNQVEGEVQIVSTAATQIFLATCINGVCWTYVHGA 60
 :|||||

Db 1026 LAPITAYAAQOTRGLGCIITSLTRDKNQVEGEVQIVSTAATQIFLATCINGVCWTYVHGA 1085
 :|||||

QY 61 GTRTIASPKGPVQMYTNVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 120
 |||||

Db 1086 GTRTIASPKGPVQMYTNVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 1145
 |||||

QY 121 GDSRGSLLSPRPISYLKSGSGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 180
 |||||

Db 1146 GDSRGSLLSPRPISYLKSGSGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 1205
 |||||

QY 181 RS 182
 ||

Db 1206 RS 1207

RESULT 4

GNWVCH
 genome polyprotein - hepatitis C virus (strain Taiwan)
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain Taiwan) (nonstruc)
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 A:Note: host Homo sapiens (man)
 C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
 C:Accession: A40244
 R:Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.
 virology 188, 102-113, 1992

RESULT 3

GNWVCH
 genome polyprotein - hepatitis C virus (strain H)
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain H) (nonstruc)
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 A:Note: host Homo sapiens (man)
 C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
 C:Accession: A36814; A41546
 R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
 submitted to GenBank, July 1992
 A:Description: Genomic structure of the human prototype strain H of hepatitis C virus
 A:Reference number: A36814
 A:Accession: A36814
 A:Molecule type: genomic RNA
 A:Residues: 1-3011 <INC>
 A:Cross-references: GB:M67463; NID:q329737; PIDN:AAA45534.1; PID:q329738
 R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
 Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
 A:Title: Genomic structure of the human prototype strain H of hepatitis C virus: com
 A:Reference number: A41546; MUID:92052256; PMID:1658800
 A:Contents: annotation
 A:Note: neither amino acid nor nucleotide sequence is given
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruc
 F:1-115/Product: capsid protein C #status predicted <GPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <WEE>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: hepatitis C virus genome
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B
 F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
 F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,234,305,325,417,433,430,448,476,532,540,556,576,623,645,1213,1255,2041,224

Query Match 97.3%; Score 927; DB 1; Length 3011;
 Best Local Similarity 96.2%; Pred. No. 1.9e-79;
 Matches 175; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 MAPITAYAAQOTRGLGCIITSLTRDKNQVEGEVQIVSTAATQIFLATCINGVCWTYVHGA 60
 :|||||

Db 1026 LAPITAYAAQOTRGLGCIITSLTRDKNQVEGEVQIVSTAATQIFLATCINGVCWTYVHGA 1085
 :|||||

QY 61 GTRTIASPKGPVQMYTNVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 120
 |||||

Db 1086 GTRTIASPKGPVQMYTNVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 1145
 |||||

QY 121 GDSRGSLLSPRPISYLKSGSGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 180
 |||||

Db 1146 GDSRGSLLSPRPISYLKSGSGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 1205
 |||||

QY 181 RS 182
 ||

Db 1206 RS 1207


```

1145
Db 1086 GSTLAGPKPIQMTNTVDQDLVGVHPAPPARGASRTPCTCGSSDLYLVTRHADVPVRRR 1145
QY 121 GDSRGSLLSPRPISYLKSGSGGPGLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 180
1146 GDSRGSLLSPRPISYLKSGSGGPGLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESMETTM 1205
QY 181 RS 182
1146 GDSRGSLLSPRPISYLKSGSGGPGLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESMETTM 1205
Db 1206 RS 1207

RESULT 7
A45573
genome polyprotein - hepatitis C virus (strain JT)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain JT)
C:Species: hepatitis C virus
A:Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: S
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
A:Accession: A45573
R:Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata,
Virus Res. 23, 39-53, 1992
A:Reference number: A45573; MUID:92295714; PMID:1318627
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-3010 <AN>
A:Cross-references: GB:D11168; GB:D01171; NID:9221612; PIDN:BA01943.1; PID:9221613
A:Experimental source: HCV-JT
A:Note: sequence extracted from NCBI backbone (NCBI:106206, NCBI:106207)
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolyase; nucleotide binding; P-loop; polyprotein; serin
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPW>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4a>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4b>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 92.8%; Score 884; DB 1; Length 3010;
Best Local Similarity 88.5%; Pred. No. 2.3e-75;
Matches 161; Conservative 15; Mismatches 6; Indels 0; Gaps 0;

QY 1 MAPITAAQOTRGLGCIITSLTGROKNQVEGEVQIVSTAAQTFLATCINGVCWTYHGA 60
1026 LAPITAAQOTRGLGCIITSLTGROKNQVEGEVQIVSTAAQTFLATCINGVCWTYHGA 1085
Db 1086 GSKTAGPKGPIQMTNTVDQDLVGVHPAPPARGASRTPCTCGSSDLYLVTRHADVPVRRR 1145
QY 61 GTRTIASPKGPIQMTNTVDQDLVGVHPAPPARGASRTPCTCGSSDLYLVTRHADVPVRRR 120
1086 GSKTAGPKGPIQMTNTVDQDLVGVHPAPPARGASRTPCTCGSSDLYLVTRHADVPVRRR 1145
Db 1086 GSKTAGPKGPIQMTNTVDQDLVGVHPAPPARGASRTPCTCGSSDLYLVTRHADVPVRRR 1145
QY 121 GDSRGSLLSPRPISYLKSGSGGPGLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 180
1146 GDSRGSLLSPRPISYLKSGSGGPGLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESMETTM 1205
Db 1206 RS 1207

RESULT 8
S18030
genome polyprotein - hepatitis C virus (isolate JKI)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (isolate JKI)
C:Species: hepatitis C virus
A:Title: The complete coding sequence of hepatitis C virus genotype 5a, the predomi
C:Reference number: JC5620; MUID:97366593; PMID:9223423

```

```

A:Variety: isolate JKI
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 23-Mar-2001
A:Accession: S18030; S33570; A48332; S18029
R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
Submitted to the EMBO Data Library, September 1991
A:Description: A whole genome of hepatitis C virus cDNA was isolated from a single p
A:Reference number: S18028
A:Accession: S18030
A:Molecule type: genomic RNA
A:Residues: 1-3010 <NON>
A:Cross-references: EMBL:X61596; NID:959478; PIDN:CAA43793.1; PID:959479
A:Experimental source: isolate JKI from an individual
R:Honda, M.; Kaneko, S.; Uonuma, M.; Kobayashi, K.; Murakami, S.
Arch. Virol. 128, 163-169, 1993
A:Title: Sequence analysis of putative structural regions of hepatitis C virus isolat
A:Reference number: A48332; MUID:93119270; PMID:8380322
A:Accession: S33570
A:Molecule type: genomic RNA
A:Residues: 1-547; 'T', 549-621, 'V', 623-624, 'S', 626-652, 'DL', 655-761, 'T', 763-782 <HOW>
A:Cross-references: EMBL:X61591
A:Note: this sequence is inconsistent with the nucleotide translation
A:Note: the authors translated the codon AGG for residue 43 as Pro, TGG for residue
as Trp, and TTC for residue 771 as Ser
A:Note: sequence extracted from NCBI backbone (NCBI:121747, NCBI:121748)
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolyase; nucleotide binding; P-loop; polyprotein; s
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPW>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4a>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4b>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate

Query Match 90.7%; Score 864; DB 1; Length 3010;
Best Local Similarity 86.8%; Pred. No. 1.9e-73;
Matches 158; Conservative 15; Mismatches 9; Indels 0; Gaps 0;

QY 1 MAPITAAQOTRGLGCIITSLTGROKNQVEGEVQIVSTAAQTFLATCINGVCWTYHGA 60
1026 LAPITAAQOTRGLGCIITSLTGROKNQVEGEVQIVSTAAQTFLATCINGVCWTYHGA 1085
Db 1026 LAPITAAQOTRGLGCIITSLTGROKNQVEGEVQIVSTAAQTFLATCINGVCWTYHGA 1085
QY 61 GTRTIASPKGPIQMTNTVDQDLVGVHPAPPARGASRTPCTCGSSDLYLVTRHADVPVRRR 120
1086 GSKTAGPKGPIQMTNTVDQDLVGVHPAPPARGASRTPCTCGSSDLYLVTRHADVPVRRR 1145
Db 1086 GSKTAGPKGPIQMTNTVDQDLVGVHPAPPARGASRTPCTCGSSDLYLVTRHADVPVRRR 1145
QY 121 GDSRGSLLSPRPISYLKSGSGGPGLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 180
1146 GDSRGSLLSPRPISYLKSGSGGPGLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESMETTM 1205
Db 181 RS 182
1206 RS 1207

RESULT 9
JC5620
genome polyprotein - hepatitis C virus (isolate EUH1480)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (isolate EUH1480)
C:Species: hepatitis C virus
A:Title: The complete coding sequence of hepatitis C virus genotype 5a, the predomi
C:Reference number: JC5620; MUID:97366593; PMID:9223423

```


A:Accession: PQ0559
A:Molecule type: mRNA
A:Residues: 2678-2729 <KAT>
A:Cross-references: GB:D10562; GB:D90518; NID:g221523; PIDN:BAA01418.1; PID:g221524
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5; status predicted <CP>
F:1-115/Product: capsid protein C #status predicted <CP>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-733/Product: nonstructural protein NS1 #status predicted <NS1>
F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>
F:1011-1619/Product: hepatitis C virus genome polyprotein
F:1234-1241/Region: nucleotide-binding motif A (P-loop)
F:1316-1321/Region: nucleotide-binding motif B
F:1320-1323/Region: DEXH motif
F:1620-1866/Product: nonstructural protein NS4a #status predicted <NS4a>
F:1867-2017/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,233,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,20,279,325,332,335,336,337,338,339,340,341,342,343,344,345,346,347,348,349,350,351,352,353,354,355,356,357,358,359,360,361,362,363,364,365,366,367,368,369,370,371,372,373,374,375,376,377,378,379,380,381,382,383,384,385,386,387,388,389,390,391,392,393,394,395,396,397,398,399,400,401,402,403,404,405,406,407,408,409,410,411,412,413,414,415,416,417,418,419,420,421,422,423,424,425,426,427,428,429,430,431,432,433,434,435,436,437,438,439,440,441,442,443,444,445,446,447,448,449,450,451,452,453,454,455,456,457,458,459,460,461,462,463,464,465,466,467,468,469,470,471,472,473,474,475,476,477,478,479,480,481,482,483,484,485,486,487,488,489,490,491,492,493,494,495,496,497,498,499,500,501,502,503,504,505,506,507,508,509,510,511,512,513,514,515,516,517,518,519,520,521,522,523,524,525,526,527,528,529,530,531,532,533,534,535,536,537,538,539,540,541,542,543,544,545,546,547,548,549,550,551,552,553,554,555,556,557,558,559,560,561,562,563,564,565,566,567,568,569,570,571,572,573,574,575,576,577,578,579,580,581,582,583,584,585,586,587,588,589,590,591,592,593,594,595,596,597,598,599,600,601,602,603,604,605,606,607,608,609,610,611,612,613,614,615,616,617,618,619,620,621,622,623,624,625,626,627,628,629,630,631,632,633,634,635,636,637,638,639,640,641,642,643,644,645,646,647,648,649,650,651,652,653,654,655,656,657,658,659,660,661,662,663,664,665,666,667,668,669,670,671,672,673,674,675,676,677,678,679,680,681,682,683,684,685,686,687,688,689,690,691,692,693,694,695,696,697,698,699,700,701,702,703,704,705,706,707,708,709,710,711,712,713,714,715,716,717,718,719,720,721,722,723,724,725,726,727,728,729,730,731,732,733,734,735,736,737,738,739,740,741,742,743,744,745,746,747,748,749,750,751,752,753,754,755,756,757,758,759,760,761,762,763,764,765,766,767,768,769,770,771,772,773,774,775,776,777,778,779,780,781,782,783,784,785,786,787,788,789,790,791,792,793,794,795,796,797,798,799,800,801,802,803,804,805,806,807,808,809,810,811,812,813,814,815,816,817,818,819,820,821,822,823,824,825,826,827,828,829,830,831,832,833,834,835,836,837,838,839,840,841,842,843,844,845,846,847,848,849,850,851,852,853,854,855,856,857,858,859,860,861,862,863,864,865,866,867,868,869,870,871,872,873,874,875,876,877,878,879,880,881,882,883,884,885,886,887,888,889,890,891,892,893,894,895,896,897,898,899,900,901,902,903,904,905,906,907,908,909,910,911,912,913,914,915,916,917,918,919,920,921,922,923,924,925,926,927,928,929,930,931,932,933,934,935,936,937,938,939,940,941,942,943,944,945,946,947,948,949,950,951,952,953,954,955,956,957,958,959,960,961,962,963,964,965,966,967,968,969,970,971,972,973,974,975,976,977,978,979,980,981,982,983,984,985,986,987,988,989,990,991,992,993,994,995,996,997,998,999,1000,1001,1002,1003,1004,1005,1006,1007,1008,1009,1010,1011,1012,1013,1014,1015,1016,1017,1018,1019,1020,1021,1022,1023,1024,1025,1026,1027,1028,1029,1030,1031,1032,1033,1034,1035,1036,1037,1038,1039,1040,1041,1042,1043,1044,1045,1046,1047,1048,1049,1050,1051,1052,1053,1054,1055,1056,1057,1058,1059,1060,1061,1062,1063,1064,1065,1066,1067,1068,1069,1070,1071,1072,1073,1074,1075,1076,1077,1078,1079,1080,1081,1082,1083,1084,1085,1086,1087,1088,1089,1090,1091,1092,1093,1094,1095,1096,1097,1098,1099,1100,1101,1102,1103,1104,1105,1106,1107,1108,1109,1110,1111,1112,1113,1114,1115,1116,1117,1118,1119,1120,1121,1122,1123,1124,1125,1126,1127,1128,1129,1130,1131,1132,1133,1134,1135,1136,1137,1138,1139,1140,1141,1142,1143,1144,1145,1146,1147,1148,1149,1150,1151,1152,1153,1154,1155,1156,1157,1158,1159,1160,1161,1162,1163,1164,1165,1166,1167,1168,1169,1170,1171,1

```

Query Match          74.7%   Score 712;   DB 1;   Length 3033;
Best Local Similarity 69.8%;   Pred. No. 5.4e-59;
Matches 127;   Conservative 29;   Mismatches 26;   Indels 0;   Gaps 0;

QY 1 MAPITAYAAQTGRLGCGITITSLTRDKNOVEGEVOIVSTAAQTFTATCINGVCWTVYHGA 60
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1030 LAPITAYAAQTGRLGIGIVVSMIGRDKTEQAGEIQVLSTVTSQFLGTTISGVLWTVYHGA 1089

QY 61 GTRTIASPKGPVQMTYNTVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVPVRRR 120
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1090 GNKTLASRGVPTQMYSSAGDVLGWPSPGPTKSLEPCTCGAVDLYLVTRNADVIPARRR 1149

QY 121 GDSRGLSLSPRISYLGKSSGGPLCPAGHAVGIFRAAVCTRCVAKAVDFIPVESLETTM 180
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1150 GDXKALLSPRLSTLKGSSGGPVLCRCHAVGVFRAAVCSRGVAKSIDFIPVETLIDVT 1209

QY 181 KS 182
   ||
Db 1210 RS 1211

RESULT 12
T08841
Polyprotein - douroucouli hepatitis GB virus A
C:Species: douroucouli hepatitis GB virus A
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000
C:Accession: T08841
J. Gen. Virol. 79, 41-45, 1998
A:Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.
A:Reference number: 216486; MCID:98120818; PMID:9460920
A:Accession: T08841
A:Status: translated from GB/EMBL/DDRJ
A:Molecule type: mRNA
A:Residues: 1-3005 <ERR>
A:Cross-references: EMBL:AF023425; NID:g2828599; PIDN:AAC40502.1; PID:g282860C
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: polyprotein

Query Match          28.1%;   Score 267.5;   DB 2;   Length 3005;
Best Local Similarity 33.1%;   Pred. No. 1e-16;
Matches 60;   Conservative 32;   Mismatches 78;   Indels 11;   Gaps 4;

QY 2 APITAYAAQTGRLGCGIITSLTRDKNOVEGEVOIVSTAAQTFLATCINGVCWTVYHAG 6;
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 979 APVVV-MORGUGLFTSVKTSMLGRDERRHSGSIVVLTCTSTRSMGTGCVNGVMYTFHGSN 1037

QY 62 TRTIASPKGPVQMTYNTVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVPVRRRG 121
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1038 ARTLAGPVPVNCRWSPSDOVAVYPLPSGASCLEPCCKGTSQVWCIRN-DGALCHGRL 1095

QY 122 DSRGSLSPRISYLGKSSGGPLCPAGHAVGIFRAAVCTRCG-----AKAVDFIPVE 174
   | |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1096 SKLVELDLPTFISDFRSGSSPILCDGHHVGNM-VSVLHRGKVTGTVRYVKWPETLPKD 1154

QY 175 S 175
   |
Db 1155 S 1155

RESULT 13
T08839
Polyprotein - marmoset hepatitis GB virus A
C:Species: marmoset hepatitis GB virus A
C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 17-Nov-2000
C:Accession: T08839
R. Erker, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.
J. Gen. Virol. 79, 41-45, 1998
A:Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.
A:Reference number: 216486; MUID:98120818; PMID:9460920
A:Accession: T08839
A:Status: translated from GB/EMBL/DBDJ
A:Molecule type: genomic RNA

A:Residues: 1-2970 <ERR>
A:Cross-references: EMBL:AF023424; NID:g2828597; PIDN:AAC40501.1; PID:g2828598
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: polyprotein

Query Match          26.8%;   Score 255.5;   DB 2;   Length 2970;
Best Local Similarity 30.1%;   Pred. No. 1.4e-15;
Matches 59;   Conservative 36;   Mismatches 68;   Indels 33;   Gaps 6;

QY 2 APITAYAAQTGRLGCGIITSLTRDKNOVEGEVOIVSTAAQTFLATCINGVCWTVYHAG 61
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 970 APVVVH-HHGKGFVGVKTSMTGDETHGVNVVLGTSTRSMGTGCVNGVMYTFHGSN 1028

QY 62 TRTIASPKGPVQMTYNTVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVPVRRRG 121
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1029 ARTLAAGPVPVNRWSSDSDVAVPLPVGAKLEPCCKQPOGVWVI-----RN 1077

QY 122 DSRGSLSL-----PRPISYLGKSSGGPLCPAGHAVGIFRAAVCTRG----- 163
   | |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1078 D--GALCHGTLGRVVELDLPAELCDFRGSSGSPILCDGHHVGNM-LSVLHRGSRVTGIR 1134

QY 164 VAKAVDFIPVESLETT 179
   | :|||:|||||:
Db 1135 YTKPWETLPREAITHT 1150

RESULT 14
H83144
Probable aromatic acid decarboxylase PA4019 [imported] - Pseudomonas aeruginosa (str.
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: H83144
R. Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.;
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic p
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: H83144
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-209 <STO>
A:Cross-references: GB:AE004818; GB:AE004091; NID:g9950200; PIDN:AAG07406.1; GSPDB:GI
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA4019
C:Superfamily: dedF protein

Query Match          9.1%;   Score 87;   DB 2;   Length 209;
Best Local Similarity 26.2%;   Pred. No. 0.69;
Matches 55;   Conservative 18;   Mismatches 61;   Indels 76;   Gaps 12;

QY 8 AQTGRLGCGIITSLTRDKNOVEGEVO-IVSTAAQTFLATCINGVCWTVYHAGTRTIA 66
   || |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 17 AQYGLRLDCLV-----QEREVHFLTSKAAQLVMAT-----ETDVA 53

QY 67 SPKGP-----VIQMTYNTVDKLVGWPAPQGSRLTPT-----CTCGSSDL 105
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 54 LPAKPAQMAQAFLEYCGAAAGQIRVEGQD-----WMAPPASGSSAPNNAVIPCSTGTL 108

QY 106 -----YLVTRHADVPVRRRGDSRGLSLSPR--PIS-----YLRGSSGGPLCLCPA 148
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 109 SAVATGACNNLIERAADVALKER---RPLVLVPREAPFSSIHLENMLKLSNLGVLTPA 164

QY 149 GHVAGIFRAAVCTRCGVAKAVDFIPVESLET 178
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 165 --APGFYHQ---POSVEDLVDFVVARILNT 189

RESULT 15
I39383
angio-associated migratory cell protein - human
C:Species: Homo sapiens (man)

```

Search completed: August 30, 2003, 19:20:24
Job time : 17.9789 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 30, 2003, 19:20:43 : Search time 1764.86 seconds
(without alignments)
2506.388 Million ccl1 updates/sec

Title: US-09-965-594-1
Perfect score: 953
Sequence: 1 MAPTAYAQTRGLGCIIT.....GVAKAVDFIPVESLETIMRS 182

Scoring table:
BLOSUMP62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-O=/cgn2.1/USPTO_spool/US09965594/runat_29082003.151919.28322/app_query.fasta_1.2672
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09965594 -RCGN_1_12630.@runat_29082003.151919.28322 -NCPU=6 -ICPU=3
-NO_WMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1.*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	106	11.1	984	10	BF304699	BF304699 601888252
C 2	103.5	10.9	1199	13	BQ892487	BQ892487 AGENCOURT
C 3	101	10.6	515	14	CA023748	CA023748 HZ47E17r
C 4	101	10.6	583	12	BM374064	BM374064 EBma03.SQ
C 5	99	10.4	615	12	BJ001625	BJ001625 BJ001625
C 6	99	10.4	643	12	BJ024121	BJ024121 BJ024121
C 7	99	10.4	754	12	BJ016176	BJ016176 BJ016176
C 8	98.5	10.3	961	10	BF203316	BF203316 601865914
C 9	98.5	10.3	1031	14	CB950999	CB950999 AGENCOURT
C 10	98.5	10.3	1141	11	AK080545	AK080545 Mus muscu
C 11	97.5	10.2	779	10	BF631437	BF631437 HVSMB001
C 12	96.5	10.1	844	11	CNS09045	AK080545 Mus muscu
C 13	96	10.1	701	10	BF863244	BF631437 HVSMB001
C 14	96	10.1	846	10	BF182274	BF863244 963042C02
C 15	95.5	10.0	901	10	BF307233	BF182274 601804028
C 16	95.5	10.0	958	10	BG420860	BF307233 601891502
C 17	95	10.0	407	9	AW785806	BG420860 602452062
C 18	95	10.0	460	14	CB883286	AW785806 117260.MA
C 19	94.5	9.9	539	28	BH349665	CB883286 HQ01M02w
C 20	94	9.9	582	14	CB286751	BH349665 CH230-65E
C 21	94	9.9	1052	10	BG398041	CB286751 CMD45_C08
C 22	94	9.9	1283	13	BQ709745	BG398041 602439571
C 23	93.5	9.8	736	12	BI768830	BQ709745 AGENCOURT
C 24	93.5	9.8	938	13	BQ894657	BI768830 603057734
C 25	93.5	9.8	993	9	AL555424	BQ894657 AGENCOURT
C 26	93	9.8	457	29	CNS02XOL	AL555424 AL555424
C 27	93	9.8	470	13	BQ758584	AL219930 Tetraodon
C 28	93	9.8	1018	12	BQ054587	BQ758584 EBma07.SQ
C 29	93	9.8	1291	10	BE622016	BQ054587 AGENCOURT
C 30	92.5	9.7	655	14	CB868789	BE622016 601440668
C 31	92.5	9.7	1008	12	BI755608	CB868789 HC09614w
C 32	92.5	9.7	1411	11	BC020343	BI755608 603027112
C 33	92	9.7	756	14	CD348815	BC020343 Homo sapi
C 34	92	9.7	832	10	BG387051	CD348815 UI-M-FY0-
C 35	92	9.7	871	10	BG178418	BG387051 602454749
C 36	92	9.7	898	29	CNS01VR5	BG178418 602330206
C 37	92	9.7	963	10	BF794182	AL169466 Tetraodon
C 38	92	9.7	1329	13	BQ960995	BF794182 602355566
C 39	92	9.7	1640	10	BF180599	BQ960995 AGENCOURT
C 40	91.5	9.6	422	14	CB763743	BF180599 601808704
C 41	91.5	9.6	539	10	BE757615	CB763743 AMGNNUC:S
C 42	91.5	9.6	641	9	AU127824	BE757615 212104.MA
C 43	91.5	9.6	691	10	BB632604	AU127824 AU127824
C 44	91.5	9.6	701	14	CD262790	BB632604 BB632604
C 45	91.5	9.6	844	12	BI198486	CD262790 psMA019xE
						BI198486 602760491

ALIGNMENTS

RESULT 1
BF304699/c
LOCUS 601888252F1 NIH_MGC_17 Homo sapiens CDNA clone IMAGE:412276 5',
DEFINITION BF304699 984 bp mRNA linear EST 21-NOV-2000
 mRNA sequence.
ACCESSION BF304699
VERSION BF304699.1 GI:11251586
KEYWORDS EST.
SOURCE Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 984)

AUTHORS NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-r@mail.nih.gov
 Tissue Procurement: AACC
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: LHCMI005 row: g column: 13
 High quality sequence stop: 646.
 Location/Qualifiers
 FEATURES
 source
 1..984
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:412276"
 /tissue_type="rhabdomyosarcoma"
 /lab_host="NIH-MGC-17"
 /clone_lib="NIH-MGC-17"
 /note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI; Site_2: XhoI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
 BASE COUNT 133 a 329 c 351 g 171 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 4.34 Length: 984
 Score: 105.00 Matches: 33
 Percent Similarity: 45.24% Conservative: 5
 Best Local Similarity: 39.29% Mismatches: 24
 Query Match: 11.12% Indels: 22
 DB: 10 Gaps: 5
 US-09-965-594-1 (1-182) x BF304699 (1-984)
 QY 86 TrpProAlaIuProGInGlySerArgSerLeuThr---ProCysThrCysGlySerSerAsp 104
 Db 646 TGGCCAGTCCAGCGCATCCCGTCCGAGAGAGGACCGTGTACCTGC----- 599
 QY 105 LeuTyrLeuValThrArgHisAlaAspValIleProValAlaArgArgGlyAspSerArg 124
 Db 598 -----ACCAGSCAGCGCAGCAACATACATGCAAGGACGCTGCT---TCCCGC 554
 QY 125 GlySerLeuLeuSerProArgPro-----IleSerTyrLeuLysGlySer 139
 Db 553 GGGCCCTCTTGTGGGGAAGACCTCGATGTGTCCAAAGCTCGCGCTGCTCTACTGGAAGT 494
 QY 140 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 159
 Db 493 CGGCAGCTCCGGTCAGGTGCAGC-----TTCAGCGCCCGGGG 455
 QY 160 CysThrArgGly 163
 Db 454 TCGCGCCGAGGA 443
 RESULT 2
 BQ892487
 LOCUS BQ892487 1199 bp mRNA linear EST 16-AUG-2002
 DEFINITION AGENCOURT_8417538 Lupski-sympathetic_trunk Homo sapiens cDNA clone
 IMAGE:6192708 5', mRNA sequence.
 ACCESSION BQ892487
 VERSION BQ892487.1 GI:22284501
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1199)
 NIH-MGC http://mgi.nci.nih.gov/
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-r@mail.nih.gov
 Tissue Procurement: Dr. James R. Lupski
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LHAM3595 row: c column: 13
 High quality sequence start: 57
 High quality sequence stop: 394.
 Location/Qualifiers
 FEATURES
 source
 1..1199
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6192708"
 /sex="male"
 /tissue_type="sympathetic trunk"
 /dev_stage="adult, 16 yr"
 /lab_host="DH10B"
 /clone_lib="Lupski-sympathetic-trunk"
 /note="Vector: pCMV-SPORT6 (Life Technologies); Site_1: NotI; Site_2: SalI; cDNA made by oligo-dT priming. Directionally cloned using the following adaptors: 5'-TCGACCCAGCGCTCCG-3' and 5'-GACTAGTTAGTAGCGGCGCCCT(15)-3'. Size selected > 1 kb for average insert length 1.9 kb. This is a primary library, non-amplified. Library constructed by Life Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor College of Medicine); available through Life Technologies."
 BASE COUNT 255 a 362 c 343 g 211 t 28 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 9.79 Length: 1199
 Score: 103.50 Matches: 41
 Percent Similarity: 37.42% Conservative: 17
 Best Local Similarity: 26.45% Mismatches: 53
 Query Match: 10.86% Indels: 44
 DB: 13 Gaps: 6
 US-09-965-594-1 (1-182) x BQ892487 (1-1199)
 QY 54 TrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLysGlyProValIle 73
 Db 484 TGGCATCCATTTTAAATAAGGGTGCTCTGTTAATCATGCGGCCACGCGCCGCTGATA 543
 QY 74 GlnMetTyrTrpAsnValAspLysAspLeuValGlyTrpProAlaProGlnGlySerArg 93
 Db 544 CTTCATTACCATGTGACGTGACTTT----- 573
 QY 94 SerLeuThrProCysThr-----CysGlySerSerAsp 104
 Db 574 ---TGTGCTGCTGCACAGACCCCATGACCGATGTTGGCTTATGTGGAACGCGAG 630
 QY 105 LeuTyr-LeuValThr-----ArgHisAlaAspValIleProValArg----- 118
 Db 631 CGCTTCATTGGCCACTCCCTTCCTATAAAACACGCCAAGCTGCTTCATGGGCGGCT 690
 QY 119 -----ArgArgGlyAspSerArgGlySerLeuLeu-- 128
 Db 691 GGGTGTGGCAGCGCGAAGCGGGTGGGCGATGTGGTAGGACTCGGGGGCGGATTCTCTG 750
 QY 129 -----SerProArgProIleSerTyrLeuLys-----GlySerSerG1 141


```

REFERENCE 1 (bases 1 to 643)
AUTHORS Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
TITLE Medaka EST Project in Takeda's lab
JOURNAL Unpublished
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES             source
     source
     1..643
     /organism="Oryzias latipes"
     /mol_type="mRNA"
     /strain="Hd-rR"
     /db_xref="taxon:8090"
     /clone="MF01SSA143D12"
     /sex="mixture of female and male"
     /tissue_type="whole embryo"
     /dev_stage="segmentation stage 20 - 25"
     /clone_lib="MF01SSA cDNA"

BASE COUNT  171 a  148 c  148 g  176 t
ORIGIN

Alignment Scores:
Pred. No.:      11.7      Length:      643
Score:          99.00     Matches:      42
Percent Similarity: 33.77%  Conservative: 9
Best Local Similarity: 27.81%  Mismatches: 50
Query Match:      10.39%    Indels:    50
DB:              12        Gaps:      7

US-09-965-594-1 (1-182) x BJ024121 (1-643)
Qy  27  LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAla 46
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  242  AAAAATGACGTAGAACCAACCAACACACATCCACACACATGTTCTGTGTTCTACGGGCT 301
Qy  47  ThrCysileAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 66
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  302  -----TGTGGAGACCTATCACAGTTCTCTGCTTTAGACACACGGCA 343
Qy  67  SerProLys-----GlyProValIleGlnMetTyrThrAsnValAspLys 81
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  344  GCTCTGGCGGGGAGGAGCTCTGGCCAGTTGTG----- 379
Qy  82  AspLeuValClyTrpProAlaProGlnGlySerArgSerLeuThrPro----- 97
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  380  -----ACTCTGGAGACGAGAGCTCCTCGCCAGTTGTG----- 418
Qy  98  -----CysThrCysGlySerAspLeuTyrLeuValThrArg----- 110
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  419  CTGACGGGATCGGATGCTGCTGCT-----TTGGTTCTCTCTCTCTCTGATCA 469
Qy  111  -----HisAlaAspValIleProValArgArgGlyAspSer 123
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  470  TCTTCTCACTGACCTTCCACATCCAGGTGTGCCAGCGCTGCTGATGAGGCTGATGG 529
Qy  124  ArgGlySerLeuLeuSerProArg-----ProIleSerTyrLeuLysGlySerSer 140
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  530  AGAGCGCGGACAGCAGCTCGGGGTGAATCTCTGCAGGACGCTTTCACGCGGATCA 589
Qy  141  GlyGlyProLeuLeuCysProAlaGlyHisAla 151
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  590  GGAGGACCGACTCGCTGCAGAGCTCTGCTGCA 622

RESULT 7
LOCUS BJ016176 754 bp mRNA linear EST 05-DEC-2001
DEFINITION BJ016176 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA025C02 3',
mRNA sequence.
ACCESSION BJ016176

```

```

VERSION BJ016176.1 GI:17376695
KEYWORDS EST.
SOURCE Oryzias latipes (Japanese medaka)
ORGANISM Oryzias latipes
REFERENCE 1 (bases 1 to 754)
AUTHORS Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
TITLE Medaka EST Project in Takeda's lab
JOURNAL Unpublished
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES             source
     source
     1..754
     /organism="Oryzias latipes"
     /mol_type="mRNA"
     /strain="Hd-rR"
     /db_xref="taxon:8090"
     /clone="MF01SSA025C02"
     /sex="mixture of female and male"
     /tissue_type="whole embryo"
     /dev_stage="segmentation stage 20 - 25"
     /clone_lib="MF01SSA cDNA"

BASE COUNT  194 a  181 c  181 g  198 t
ORIGIN

Alignment Scores:
Pred. No.:      14.4      Length:      754
Score:          99.00     Matches:      42
Percent Similarity: 33.77%  Conservative: 9
Best Local Similarity: 27.81%  Mismatches: 50
Query Match:      10.39%    Indels:    50
DB:              12        Gaps:      7

US-09-965-594-1 (1-182) x BJ016176 (1-754)
Qy  27  LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAla 46
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  242  AAAAATGACGTAGAACCAACCAACACACATCCACACACATGTTCTGTGTTCTACGGGCT 301
Qy  47  ThrCysileAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 66
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  302  -----TGTGGAGACCTATCACAGTTCTCTGCTTTAGACACACGGCA 343
Qy  67  SerProLys-----GlyProValIleGlnMetTyrThrAsnValAspLys 81
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  344  GCTCTGGCGGGGAGGAGCTCTGGCCAGTTGTG----- 379
Qy  82  AspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrPro----- 97
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  380  -----ACTCTGGAGGAGGAGAGCTCCTCGCCAGTTGTG----- 418
Qy  98  -----CysThrCysGlySerAspLeuTyrLeuValThrArg----- 110
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  419  CTGACGGGATCGGATGCTGCTGCT-----TTGGTTCTCTCTCTCTGATCA 469
Qy  111  -----HisAlaAspValIleProValArgArgGlyAspSer 123
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  470  TCTTCTCACTGACCTTCCACATCCAGGTGTGCCAGCGCTGCTGATGAGGCTGATGG 529
Qy  124  ArgGlySerLeuLeuSerProArg-----ProIleSerTyrLeuLysGlySerSer 140
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  530  AGAGCGCGGACAGCAGCTCGGGGTGAATCTCTGCAGGACGCTTTCACGCGGATCA 589
Qy  141  GlyGlyProLeuLeuCysProAlaGlyHisAla 151
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```



```

Db      452 TTG---GGCACACTGTGTGTCGTGGCACAT-----ATCGATCGCCCT 490
QY      69  LysGlyProValIleGlnMetTyrThrAsnValAspLysAspLeuValClyTrpProAla 88
Db      491 AAGAGGCCCTTTACAAAACACT-----AACCTCCCTGCGCTGGCGCTGCC 535
QY      89  Pro-----GlnGlySerArgSerLeuThrProCysThrCysGly 101
Db      536 AT-GTTGGGCAAAAGAACCGTTTGTGGGTTCGGCCCTTCGCCCCCGCCCAATTGGGA 594
QY      102 SerSerAspLeuTyrLeuValThrArgHisAlaAsp-VaiIleProValArgArgGly 121
Db      595 ACCAGTGGC-----ACCACCATGGGGCTGTGTGTTCGCCCGCTTCGCCGTGG 642
QY      121 YAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGly 141
Db      643 GCAATTACAAACACCCCTTAACCGTCCCTCCCAACATATTCTTCAGCGCTCTCGA 702
QY      141 y-----GlyProLeuLeuCysProAlaGlyHisAlaValGly 153
Db      703 TTTCCTTAAGTCCCGCCCTTTGTGTACCCCGACACCATTTGTGGGA 748

RESULT 10
AK080545
LOCUS
DEFINITION
Mus musculus 7 days neonate cerebellum cDNA, RIKEN full-length
enriched library, clone:A730082L10 product:weakly similar to zinc
finger protein (fragment) [Mus musculus], full insert sequence.
AK080545
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
1
AUTHORS
Carninci, P. and Hayashizaki, Y.
TITLE
High-efficiency full-length cDNA cloning
JOURNAL
MEDLINE
99279253
PUBMED
10349636
REFERENCE
2
AUTHORS
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL
MEDLINE
20499374
PUBMED
11042159
REFERENCE
3
AUTHORS
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsumai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,
Yoneda, Y., Ishikawa, I., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL
MEDLINE
20530913
PUBMED
11076861
REFERENCE
4
AUTHORS
Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I.,
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G.,
Quackenbush, J., Schriml, L. M., Staubli, F., Suzuki, R., Tomita, M.,
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,

```

```

Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C.,
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,
Marrionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,
Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schonbach, C., Seva, T., Shibata, Y., Storch, K. F., Suzuki, H.,
Toyooka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L., S.
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohlsuki, S.
and Hayashizaki, Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
21085660
PUBMED
11217851
5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase 1 & II team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 1141)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/
location/Qualifiers
1. ..1141
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM,DB:A730082L10"
/db_xref="taxon,10090"
/clone="A730082L10"
/tissue_type="cerebellum"
/collection="RIKEN full-length enriched mouse cDNA library"
/dev_stage="7 days neonate"
<1..587
/note="unnamed protein product: putative
weakly similar to zinc finger protein (fragment) [Mus
musculus] (PIR)I48722, evidence: FASTV, 50.7%ID,
57.6%length, match=601"
/codon_start=3
/db_xref="GI:26348601"
/translation="DSCLPAAQSGRSLTPRGDGFLEKLSAARAVGPGSVAFGV
TRYGAAQPGQRRVGCARRSEGLCSRPRRQHPVPPVPHVYGLSGRRIPPPAGE
AQAAGAAQPPHPGRPHGTVPVPOGAAGLLPALAARQVPPVGRGPRAPRHS
PKPVPALGFGCGGPPAPPELLAPANGRSVGLA"
1118..1123
/note="putative"
polya_signal

```

```

polyA_site      1141      /note="putative"
BASE COUNT      244 a      316 c      353 g      228 t
ORIGIN
Alignment Scores:
Pred. No.:      27-7      Length:      1141
Score:          98.50      Matches:      51
Percent Similarity: 34.83%      Conservative: 11
Best Local Similarity: 28.65%      Mismatches: 54
Query Match:      10.34%      Indels:      57
DB:             11      Gaps:      10

US-09-965-594-1 (1-182) x AK080545 (1-1141)
QY 13 GlyLeuLeuGlyCysIleIleThrSerLeuThrGlyArgAspLysAsnGlnValGluGly 32
DB 267 GGCCTCTCCGGT-----GGCAGAGCATCCCGCCGCCGGCGGA 305
QY 33 GluValGlnIleValSerThrAlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyVal 52
DB 306 GAAGCGCAGGAGCTGGGCGCGCCGCGAG----- 335
QY 53 CysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLysGlyProVal 72
DB 336 -----CAAGTCCGCATCCCGCTGGG-----CGGCCACATGGAACAGGIC 374
QY 73 IleGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpProAlaProGlnGlySer 92
DB 375 GTG-----CCTCTCAAGGAGCG 392
QY 93 ArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeuValThrArgHisAla 112
DB 393 GCTGGCTTCTCCCTGCA-----CTCGCAGCTCGCAAGTT 428
QY 113 AspValIleProValArg---ArgArgLysAspSerArgGlySerLeuLeuSerProArg 131
DB 429 CTGTGTGACCGGTAGCGCGGAGGACCAAGAGGCGCAGGACACACACGCCCAAG 488
QY 132 ProIle-----SerTyrLeuLysGlySerSerGlyGlyProLeuLeu 145
DB 489 CCGGTCTCTACAGCTTGGCTTTCGTTTGGCAGGCTGGCTCTCTCCCTCCCTCA 548
QY 146 CysProAla---GlyHisAlaValAlaGlyPheArgAlaAlaValCysThrArgGlyVal 164
DB 549 GCCCGCGGATGAGGAGGCTGTAGGGTTG---GCCCTGTAAAGTGTGTGAGTCCGGAGAC 605
QY 165 AlaIysAlaVal-----AspPheIleProValGluSerLeuGluThrThrMet 180
DB 606 TTGAGGGTGGTTGGCTCTGAGCTGACACCAAGCTCTCTCTGCGAGGTACACTT 659

RESULT 11
BF631437
LOCUS          779 bp      mRNA      linear      EST 22-OCT-2001
DEFINITION    HVSMED0015P05f Hordeum vulgare seedling shoot EST library
              clone HVSMED0015P05f. Hordeum vulgare subsp. vulgare cDNA
              sequence.
ACCESSION     BF631437
VERSION       BF631437.2 GI:13092107
KEYWORDS
SOURCE        Hordeum vulgare subsp. vulgare
ORGANISM      Hordeum vulgare subsp. vulgare
              Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
              Triticeae; Hordeum.
REFERENCE     1 (bases 1 to 779)
AUTHORS      Wing,R., Close,T.J., Kleinbols,A., Wise,R., Begum,D., Frisch,D., Yu
              ,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W., Fenton
              ,R.D., Gates,R. and Main,D.
TITLE        Development of a genetically and physically anchored EST resource
              for barley genomics: Morex drought-stressed seedling shoot cDNA
              library
JOURNAL       Unpublished

```

```

COMMENT
On Dec 19, 2000 this sequence version replaced gi:11895595.
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hq bases = 455
Seq primer: AATTACCTCTCACTAAAGGG
High quality sequence stop: 588.
Location/Qualifiers
1..779
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Morex"
/db_xref="taxon:112509"
/clone="HVSMED0015P05f"
/tissue_type="Seedling shoot"
/lab_host="TJC121"
/clone_lib="Hordeum vulgare seedling shoot EST library
HVCNA0002 (Dehydration stress)"
/note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI;
Seeds were surface sterilized then germinated under axenic
conditions in the dark at room temperature on filter paper
with water, nystatin and cefotaxime in covered
crystallization dishes. Five-day old seedlings were
incubated at 90% RH for 24 hr. Shoots were then harvested,
total RNA was prepared, poly(A) RNA was purified, one
primary unamplified cDNA library was made, 600000 pfu were
in vivo excised to give pBluescript SK(-) cDNA phagemids.
These steps were performed in the TJ Close laboratory at
the University of California, Riverside (Choi, Close,
Fenton). Phagemids were plated and picked at the Clemson
University Genomics Institute (CUGI) (Begum, Palmer,
Frisch, Atkins and Wing). Plasmid DNA preparations, DNA
sequencing and sequence analysis were performed at CUGI
(Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).
The sequence has been trimmed to remove vector sequence
and contains a minimum of 100 bases of phred value 20 or
above. For more details on library preparation and
sequence analysis see
http://www.genome.clemson.edu/projects/barley. To order
this clone see http://www.genome.clemson.edu/orders Also
see Close TJ, Wing R, Kleinbols A, Wise R (2001)
Genetically and physically anchored EST resources for
barley genomics. Barley Genetics Newsletter 31:29-30.
(http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)*
BASE COUNT      147 a      216 c      311 g      105 t
ORIGIN
Alignment Scores:
Pred. No.:      21      Length:      779
Score:          97.50      Matches:      47
Percent Similarity: 39.87%      Conservative: 14
Best Local Similarity: 30.72%      Mismatches: 63
Query Match:      10.23%      Indels:      30
DB:             10      Gaps:      6

US-09-965-594-1 (1-182) x BF631437 (1-779)
QY 35 GlnIleValSerThrAlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrp 54
DB 1 CAGTGGCAACGGCGGCCCGCGGAACACTACAGGACGGTGC-----TGTGG 48
QY 55 ThrValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLysGlyProValIle--- 73
DB 49 ACCCGGTGGAGGGCGGTGGCGCCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 108
QY 74 -----GlnMetTyrThrAsnValAspLysAspLeuValGlyTrpPro 87
DB 109 CCCTGTTTGGCAGGTGGCCCGGTGGAGCGGCGGCGGAGGAGCGGCTGCTGCTGCTG 168

```

623	Db	ATCTTTACCGGTGCGTGCACACTCTTCTC	ACAGAGCTGGCTGCATCGCATCAGAACGATA	564
66	QY	AlaSerProLys	-----GlyProValIleleuMetTyrIhr	77
563	Db	CCCGGTTCAAGAAGCTGCTGATGCTGCTGGCGAGCGGGCCCTCTTTGCGC	-----	513
78	QY	AsnValAspLysAspLeuValGlyTTPProAlaProGlnGlySerArgLeuThrPro	97	
512	Db	-----	-----CGGTCAATTACCGCCC	498
98	QY	CysThrCys	-----GlySerSerAspLeuTyrLeuValIhrArgHisAlaAsp	113
497	Db	TGTGCTGTACTTTGTTCTCCGCTCGAA	-----ACAAACACACAGC	456
114	QY	VallIleProValArgArgArgGlyAspSerArgGlySerLeuLeuSerProArgProIle	133	
455	Db	-----	-----ACGAGCGCTCGGCAATCTGTTGCCACCATCATCGGATTCTGCTGCAT	420
134	QY	SerTyrIleuLysGlySerSerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGly	153	
419	Db	GCAATTCTGC	-----GGGCAACTGGAAGTGCCTTACCCTGCCCGTATTTCATCACTTCAACT	363
154	QY	IlePheArgAlaAlaValCysThrArgGlyValAlaLys	-----AlaValAsp	169
362	Db	ACATTCCCGCGA	-----TGTACCGGTGCTGTTTGGCAACATCATCGGATTCTGCTGCAT	309
170	QY	PheIleProVal	173	
308	Db	GGTGTTCTGTC	297	
RESULT 13				
BF863244				
LOCUS				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
COMMENT				
FEATURES				
source				

phuescript II SK- plasmids were excised from the lambda ZAP clones by superinfection with EXAssist (Stratagene) in Bernaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT 173 a 213 c 175 q 140 t

ORIGIN

Alignment Scores:

Pred. No.: 25.4 Length: 701
Score: 96.00 Matches: 32
Percent Similarity: 40.71% Conservative: 14
Best Local Similarity: 28.32% Mismatches: 45
Query Match: 10.07% Indels: 22
DB: 10 Gaps: 4

US-09-965-594-1 (1-182) x BF683244 (1-701)

Qy 57 TyrHisGlyAlaGlyThrArgThrileAlaSerProLys-----GlyProVal 72
Db 171 CACCACCATACCTTGTCTCAGGTTCCTCACACCAAAATATGCCCATACGGCCACIA 230
Qy 73 IleGlnMetThrAsnValAspLysAspLeuValGlyTrpProAlaProGlnGlySer 92
Db 231 ACAAGTTACATACCG-----AAGGACACCGCGCTGGCCACCCCTTGGAGCGG 284
Qy 93 ArgSerLeuThrProCysThrCysGlySerAspLeuValThrArgHisAla 112
Db 285 AGAAGCCGACCGTGTCTCTGGGTCTATCGCATCTCTCAATCTCCGCTATCAG 344
Qy 113 AspValle-----ProValArgArgArgGlyAspSerArg----- 124
Db 345 GAGATCATTTGTCATGTGGCTTTAGTACCCCAAGAGAGCGCTGGGAGTGGGCTTTATAA 404
Qy 125 -----GlySerLeuSerProArgProLysSerProLys 136
Db 405 GAAGCGGACGGGAATTCGGTTTCGGAAGAGTGCAGCGCCCAAGGTCACCAAGTGCTA 464
Qy 137 LysGlySerGlyGlyProLeuLeuCysProAlaGly 149
Db 465 CTCACGACGACAAATGGAGCCCTTCGGGTGTGGCGGT 503

RESULT 14
BF182274/c 846 bp mRNA linear EST 31-OCT-2000
LOCUS 601804028F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:4035102 5',
DEFINITION mRNA sequence.

ACCESSION BF182274.1 GI:11050416

VERSION EST.

KEYWORDS Mus musculus (house mouse)

SOURCE

ORGANISM

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 846)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-f@mail.nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L1AM9308 row: g column: 07

High quality sequence stop: 696.

Location/Qualifiers

FEATURES

source

1..846

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"
/clone="IMAGE:4035102"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Mam5"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: SalI;
Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Robin Humphreys,
NIH"

BASE COUNT 176 a 218 c 241 g 210 t 1 others

ORIGIN

Alignment Scores:

Pred. No.: 32.5 Length: 846
Score: 96.00 Matches: 46
Percent Similarity: 47.20% Conservative: 13
Best Local Similarity: 36.80% Mismatches: 36
Query Match: 10.07% Indels: 31
DB: 10 Gaps: 9

US-09-965-594-1 (1-182) x BF182274 (1-846)

Qy 59 GlyAlaGlyThrArg-ThrileAlaSerProLysGlyProValIleGlnMetThrAs 78
Db 757 GGTTCCTCTACCAAAACGCTGGATGAAGAAAGACCA-----CATCCTTC 710
Qy 78 nValAspLysAspLeuValGlyTrp-----ProAlaProGln 90
Db 709 GGTTCTTCTACCAAGTGGGCTGGAGAAAGTGAACACAGACAGACGCTCCCTCA 650
Qy 90 nGlySerArg-----SerLeuThrProCysThrCysGlySerSerAspLeuThr 108
Db 649 GTCCCCACGCTCTAGTAGTCTAGACAAAGTGTCTGCTGGA----- 610
Qy 108 lThrArgHisAlaAspValleProValArgArgGlyAspSerArgGlySerLeu 128
Db 609 -ACTAGACACACCT--GTAATCCAGGAGAACGCTGGAGGACACAGAGGACTCC---CT 556
Qy 128 userProArgProLysSerThrLeuLysGlySerSerGlyGlyProLeu---Leu 147
Db 555 GACCCACCTCCC---TCCGTCTCTAGCGGACCTCTCTCGGCCCCACCTCTCTGTC 499
Qy 147 oAlaGlyHis---AlaValGlyIlePheArg-----AlaAlaValCysThrAr 162
Db 498 TAGTGGGCACCTCTCCCGGACGACGACAGACTGTACTCCCTTTGGCCCTCTGCACTCT 439
Qy 162 gGlyValAlaLys 166
Db 438 TGGGATGACTGAG 426

RESULT 15

BF307233

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 901)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-f@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: L1CM1044 row: c column: 02
 High quality sequence start: 6
 High quality sequence stop: 684.
 Location/Qualifiers
 1. .901
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4137145"
 /tissue_type="rhabdomyosarcoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_17"
 /note="Organ: muscle; Vector: pOT87; Site: 1: PcoRI;
 Site 2: XhoI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCACGAG(G). Size-selected >500bp
 for average insert size 1.8kb. Library constructed by
 Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 144 a 267 c 329 g 161 t

ORIGIN

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
52	39.5	901	37
620	95.50		
Percent Similarity:	38.46%	Conservative:	8
Best Local Similarity:	31.62%	Mismatches:	28
Query Match:	10.02%	Indels:	45
DB:	10	Caps:	5

US-09-965-594-1 (1-182) x BF307233 (1-901)

Qy	52	ValCysTrpThr-VaITyHis-----GlyAlaGlyThrArgThrIleAl	66
Db	620	ATGTGCTGGAGCGGTCCCGCACGCCATCCTGAGCGGGGTCCGGCACACACACAGCGTGG	679
Qy	66	aSerProLysGlyProValIleGlnMetTyrThrAsnValAspLysAspLeuValGlyTr	86
Db	680	TGGCAGCGGTGGAGTGTGCA-----GTGTCAGGATG	715
Qy	86	pProAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTy	106
Db	716	GCOCGCCCATCCGG-----	731
Qy	106	rLeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySe	126
Db	732	-----GTGAGCCTTCGTCTCAGGGCGGTTCGGCGGGGTTTC	766
Qy	126	rLeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCy	146
Db	767	CCITTTGGGTCTGA-----CGGGCATCTCTCCAGGGCGCGTGGACTG	810
Qy	146	sProAlaGlyHisAlaValGly-----IlePheArgAlaAlaValCys	160
Db	811	TCCGGCGGTTCGGCGCGGCCGCCACACAGGTCGGCGGGCGGTGTGC	859

Search completed: August 31, 2003, 04:27:18
 Job time : 1773.86 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: August 30, 2003, 17:42:58 : Search time 44.1697 seconds
(without alignments)
700.745 Million cell updates/sec

Title: us-09-965-594-12
Perfect score: 1021
Sequence: 1 MKKGSVIVGRIVLNAYV.....VAKAVDFIPVESLETMRSP 195

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A.Geneseq_19Jun03:*

1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*

2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*

3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*

4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*

5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*

6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*

7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*

8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*

9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*

10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*

11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*

12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*

13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*

14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*

15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*

16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*

17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*

18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*

19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*

20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*

21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*

22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1021	100.0	195	AA15220	Hepatitis C virus
2	998	97.7	197	AA15221	Hepatitis C virus
3	982	96.2	195	AA15212	Hepatitis C virus
4	981	96.1	197	AA15222	Hepatitis C virus
5	959	93.9	197	AA15226	Hepatitis C virus
6	951	93.1	197	AA15223	Hepatitis C virus
7	939	92.0	197	AA15224	Hepatitis C virus
8	929	91.0	197	AA15225	Hepatitis C virus
9	899.5	88.1	191	AA144728	Hepatitis C virus

10	898.5	88.0	2816	14	AA134009	HCV-1 polyprotein.
11	895.5	87.7	3011	14	AA140120	HCV genomic amino
12	894.5	87.6	1766	10	AA192041	Sequence encoded 1
13	894.5	87.6	1786	10	AA190158	Protein sequence o
14	894.5	87.6	2261	10	AA190164	Peptide encoded by
15	894.5	87.6	2301	10	AA192047	Sequence encoded 1
16	894.5	87.6	2436	10	AA192050	Sequence encoded 1
17	894.5	87.6	2436	10	AA192050	Sequence encoded 1
18	894.5	87.6	2772	21	AA18540	Peptide encoded by
19	894.5	87.6	2894	16	AA170230	Protein encoded by
20	894.5	87.6	2955	20	AA14975	Composite hepatitis
21	894.5	87.6	2955	21	AA18541	Amino acid sequenc
22	894.5	87.6	3011	13	AA121519	Polyprotein encode
23	894.5	87.6	3011	14	AA131621	Compiled HCV sequ
24	894.5	87.6	3011	17	AA190931	Hepatitis C virus
25	894.5	87.6	3011	18	AA134480	Hepatitis C virus
26	894.5	87.6	3011	19	AA140038	HCV polyprotein.
27	894.5	87.6	3011	23	AA122049	Hepatitis C virus
28	894.5	87.6	3011	23	AA184597	Hepatitis C virus
29	894	87.6	182	21	AA15211	Hepatitis C virus
30	894	87.6	609	15	AA151170	Hepatitis C virus
31	894	87.6	631	18	AA131884	A nonstructural pr
32	894	87.6	686	23	AA18689	HCV-1 NS3/4a mutan
33	894	87.6	686	23	AA178377	Hepatitis C virus
34	894	87.6	686	24	ABG72261	Hepatitis C virus
35	893.5	87.5	3011	15	AA166995	HCV-1 NS3/4a confo
36	892	87.4	631	20	AA193482	Hepatitis C virus
37	891.5	87.3	665	20	AA124943	HCV NS3 protein.
38	891.5	87.3	2435	13	AA125135	HCV polypeptide 1.
39	891.5	87.3	2436	13	AA128582	HCV amino acid seq
40	890.5	87.2	2772	11	AA108123	Hepatitis C virus
41	890	87.2	532	23	AA121847	Hepatitis C virus
42	890	87.2	632	23	AA119905	Hepatitis C virus
43	890	87.2	686	23	AA121837	Hepatitis C virus
44	890	87.2	686	23	AA121838	Hepatitis C virus
45	890	87.2	686	23	AA121839	Hepatitis C virus

ALIGNMENTS

RESULT 1
AA15220
ID AA15220 standard; protein; 195 AA.
AC AA15220;
DT 19-DEC-2000 (first entry)
XX
XX Hepatitis C virus NS4A-NS3 fusion protease #2.
DE
XX Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW liver failure; liver cancer; mutant; mutein.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
XX
PN WO200040707-A1.
XX
PD 13-JUL-2000.
XX
PF 06-JAN-2000; 2000WO-US00345.
XX
PR 08-JAN-1999; 99US-0115271.
XX
XX (BRIM) BRISTOL-MYERS SQUIBB CO.
PA
XX
XX
PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX
XX WPI; 2000-465976/40.
DR
DR N-PSDB; AAA73329.
XX
XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1

substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic amino acid, useful for screening inhibitors that may treat hepatitis C

Claim 23: Fig 12: 66pp: English.

The present sequence is a mutated version of a fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These proteins are both essential for the replication of the virus, acting to cleave its replicative proteins from the polyprotein produced from the HCV genome. Inhibitors of the two proteins should be effective as antiviral treatments of HCV infection. This is useful as HCV can lead to chronic liver disease such as cirrhosis, liver failure and liver cancer. The present invention concerns a number of NS3 mutants and NS3-NS4A fusion proteins which can be used to identify inhibitors of this type, as well as enabling structural studies of the protease and protease:inhibitor complexes. This sequence contains the alpha-helix0-1 variant.

Sequence 195 AA:

RESULT 2	
AA15221	AA15221 standard; protein; 197 AA.
XX	
XX	AA15221:
XX	
XX	19-DEC-2000 (first entry)
XX	
DE	Hepatitis C virus NS4A-NS3 fusion protease #3.
XX	
XX	Hepatitis: NS3 protease; viral replication; chronic liver disease:
KW	liver failure; liver cancer; mutant; mutcin.
XX	
XX	Hepatitis C virus.
OS	Synthetic.
XX	
PN	MO200040707-A1.
XX	
PD	13-JUL-2000.
XX	
PF	06-JAN-2000; 2000WO-0500345.
XX	
PR	08-JAN-1999; 99US-0115271.
XX	
PA	{BRIM } BRISTOL-MYERS SQUIBB CO.
XX	
PI	Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX	
DR	WPI; 2000-465976/40.
XX	
DR	N-PSDB; AAA73330.
XX	

Modified hepatitis C virus (HCV) NS3 protease comprising at least 1 substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic amino acid, useful for screening inhibitors that may treat hepatitis C

Claim 23; Fig 13; 66pp; English.

The present sequence is a mutated version of a fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These proteins are both essential for the replication of the virus, acting to cleave its replicative proteins from the polyprotein produced from the HCV genome. Inhibitors of the two proteins should be effective as antiviral treatments of HCV infection. This is useful as HCV can lead to chronic liver disease such as cirrhosis, liver failure and liver cancer. The present invention concerns a number of NS3 mutants and NS3-NS4A fusion proteins which can be used to identify inhibitors of this type, as well as enabling structural studies of the protease and protease:inhibitor complexes. This sequence contains the alpha-helix0-1 variant.

Sequence 197 AA:

RESULT 3	
AAAB15212	Hepatitis C virus NS4A-NS3 fusion protease #1.
ID	AAAB15212 standard; protein; 195 AA.
XX	
AC	AAAB15212;
XX	
DT	19-DEC-2000 (first entry)
XX	
DE	Hepatitis C virus NS4A-NS3 fusion protease #1.
XX	
KW	Hepatitis; NS3 protease; viral replication; chn
XX	liver failure; liver cancer.
XX	
OS	Hepatitis C virus.
GS	Synthetic.
XX	
PN	WO2000040707-A1.
XX	
PD	13-JUL-2000.
XX	
PF	06-JAN-2000; 2000WO-US00345.
XX	
PR	08-JAN-1999; 99US-0115271.
XX	
PA	(BRIM) BRISTOL-MYERS SQUIBB CO.
XX	
PI	Wittekind M, Weinheimer S, Zhang Y, Goldfarb
XX	
DR	WPI; 2000-465976/40.
DR	N-PSDB; AAA73328.


```

XX
PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
PT amino acid, useful for screening inhibitors that may treat hepatitis C
PT
XX
PS Example 2; Fig 10; 66pp; English.
XX
CC The present sequence is a fusion protein created using the Hepatitis C
CC virus (HCV) NS3 and NS4A protease enzymes. These proteins are both
CC essential for the replication of the virus, acting to cleave its
CC replicative proteins from the polyprotein produced from the HCV genome.
CC Inhibitors of the two proteins should be effective as antiviral
CC treatments of HCV infection. This is useful as HCV can lead to chronic
CC liver disease such as cirrhosis, liver failure and liver cancer. The
CC present invention concerns a number of NS3 mutants and NS3-NS4A fusion
CC proteins which can be used to identify inhibitors of this type, as well
CC as enabling structural studies of the protease and protease:inhibitor
CC complexes.
XX
SQ Sequence 195 AA;
Query Match 96.2%; Score 982; DB 21; Length 195;
Best Local Similarity 97.4%; Pred. No. 3.2e-94;
Matches 190; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 MKKKGSVVIVGRVILNCAVAQOTRGEGCQETSGTGRDKNOVEGEVOIVSTAAQTFLATC 60
DB 1 MKKKGSVVIVGRVILNCAVAQOTRGLIGCIITSLTGRDKNOVEGEVOIVSTAAQTFLATC 60
QY 61 INGVCMVTYVHGAGTRTIASPKGPVIOMYTNVDKDLVGPAPOGSRLTPTCTCGSSDLY 120
DB 61 INGVCMVTYVHGAGTRTIASPKGPVIOMYTNVDKDLVGPAPOGSRLTPTCTCGSSDLY 120
QY 121 TRHADVIPVRRRGRSGSLSPRPISYLGSGGGLPCPAGHAVGIFRAAVCTRGVAK 180
DB 121 TRHADVIPVRRRGRSGSLSPRPISYLGSGGGLPCPAGHAVGIFRAAVCTRGVAK 180
QY 181 DFIPVESLETTMRSP 195
DB 181 DFIPVESLETTMRSP 195
RESULT 4
AAB15222
ID AAB15222 standard; protein: 197 AA.
XX
AC AAB15222;
XX
DT 19-DEC-2000 (first entry)
XX
DE Hepatitis C virus NS4A-NS3 fusion protease #4.
XX
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW liver failure; liver cancer; mutant; muteln.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
PN WO200040707-A1.
XX
PD 13-JUL-2000.
XX
PF 06-JAN-2000; 2000WO-US00345.
XX
PR 08-JAN-1999; 99US-0115271.
XX
PA (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX
DR WPI; 2000-465976/40.
N-PSDB; AAA73331.

```

```

XX
PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
PT amino acid, useful for screening inhibitors that may treat hepatitis C
PT
XX
PS Claim 23; Fig 14; 66pp; English.
XX
CC The present sequence is a mutated version of a fusion protein created
CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
CC proteins are both essential for the replication of the virus, acting to
CC cleave its replicative proteins from the polyprotein produced from the
CC HCV genome. Inhibitors of the two proteins should be effective as
CC antiviral treatments of HCV infection. This is useful as HCV can lead to
CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
CC The present invention concerns a number of NS3 mutants and NS3-NS4A
CC fusion proteins which can be used to identify inhibitors of this type, as
CC well as enabling structural studies of the protease and
CC protease:inhibitor complexes. This sequence contains the alpha-helix0-1
CC variant.
XX
SQ Sequence 197 AA;
Query Match 96.1%; Score 981; DB 21; Length 197;
Best Local Similarity 96.4%; Pred. No. 4.1e-94;
Matches 190; Conservative 1; Mismatches 4; Indels 2; Gaps 1;
QY 1 MKKKGSVVIVGRVILNCAVAQOTRGEGCQETSGTGRDKNOVEGEVOIVSTAAQTFLA 58
DB 1 MKKKGSVVIVGRVILNCAVAQOTRGEGCQETSGTGRDKNOVEGEVOIVSTAAQTFLA 60
QY 59 TCINGVCMVTYVHGAGTRTIASPKGPVIOMYTNVDKDLVGPAPOGSRLTPTCTCGSSDLY 118
DB 61 TCINGVCMVTYVHGAGTRTIASPKGPVIOMYTNVDKDLVGPAPOGSRLTPTCTCGSSDLY 120
QY 119 LVTRHADVIPVRRRGRSGSLSPRPISYLGSGGGLPCPAGHAVGIFRAAVCTRGVAK 178
DB 121 LVTRHADVIPVRRRGRSGSLSPRPISYLGSGGGLPCPAGHAVGIFRAAVCTRGVAK 180
QY 179 AVDFIPVESLETTMRSP 195
DB 181 AVDFIPVESLETTMRSP 197
RESULT 5
AAB15226
ID AAB15226 standard; protein: 197 AA.
XX
AC AAB15226;
XX
DT 19-DEC-2000 (first entry)
XX
DE Hepatitis C virus NS4A-NS3 fusion protease #8.
XX
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW liver failure; liver cancer; mutant; muteln.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
PN WO200040707-A1.
XX
PD 13-JUL-2000.
XX
PF 06-JAN-2000; 2000WO-US00345.
XX
PR 08-JAN-1999; 99US-0115271.
XX
PA (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX
DR WPI; 2000-465976/40.

```

DR N-PSDB; AAA73335.

XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1

PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic

PT amino acid, useful for screening inhibitors that may treat hepatitis C

PT

XX

XX Example 5; Fig 18; 66pp; English.

PS

XX The present sequence is a mutated version of a fusion protein created

CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These

CC proteins are both essential for the replication of the virus, acting to

CC cleave its replicative proteins from the polypeptide produced from the

CC HCV genome. Inhibitors of the two proteins should be effective as

CC antiviral treatments of HCV infection. This is useful as HCV can lead to

CC chronic liver disease such as cirrhosis, liver failure and liver cancer.

CC The present invention concerns a number of NS3 mutants and NS3-NS4A

CC fusion proteins which can be used to identify inhibitors of this type, as

CC well as enabling structural studies of the protease and

CC protease-inhibitor complexes. This sequence contains the alpha-helix0

CC wild-type sequence.

XX

XX Sequence 197 AA;

SQ

Query Match 93.9%; Score 959; DB 2; Length 197;

Best Local Similarity 95.4%; Pred. No. 8.2e-92;

Matches 188; Conservative 1; Mismatches 6; Indels 2; Gaps 1;

QY 1 MKKGSVVIVGRVING--AYAQTRGEGGCOETSGTRDKNOVEGEVQIVSTAAQTFLA 58

DB 1 MKKGSVVIVGRVINGSGDTAYAQTRGEGGCOETSGTRDKNOVEGEVQIVSTAAQTFLA 60

QY 59 TCINGVCVTVYHGACTRTIASPKGPVQIOMYTNVDKDLVGNPAPQGSRSLSLTPTCTGSSDLY 118

DB 61 TCINGVCVTVYHGACTRTIASPKGPVQIOMYTNVDKDLVGNPAPQGSRSLSLTPTCTGSSDLY 120

QY 119 LVTRHADVIPVRRGDSRGLSPRISYLGSSGGPCLCPAGHAGVIFRAAVCTRGVAK 178

DB 121 LVTRHADVIPVRRGDSRGLSPRISYLGSSGGPCLCPAGHAGVIFRAAVCTRGVAK 180

QY 179 AVDFIPVESLETTMRSP 195

DB 181 AVDFIPVESLETTMRSP 197

RESULT 6

AAB15223

ID AAB15223 standard; protein; 197 AA.

XX

AC AAB15223;

XX

DT 19-DEC-2000 (first entry)

XX

DE Hepatitis C virus NS4A-NS3 fusion protease #5.

XX

XX Hepatitis; NS3 protease; viral replication; chronic liver disease;

KW liver failure; liver cancer; mutant; mutein.

XX

OS Hepatitis C virus.

OS Synthetic.

XX

XX WO2000040707-A1.

PN

XX

PD 13-JUL-2000.

XX

XX 06-JAN-2000; 2000WO-US00345.

PF

XX

PR 08-JAN-1999; 99US-0115271.

XX

PA (BRIM) BRISTOL-MYERS SQUIBB CO.

XX

XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;

XX

DR WPI: 2000-465976/40.

DR N-PSDB; AAA73332.

XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1

PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic

PT amino acid, useful for screening inhibitors that may treat hepatitis C

PT

XX

XX Claim 23; Fig 15; 66pp; English.

PS

XX The present sequence is a mutated version of a fusion protein created

CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These

CC proteins are both essential for the replication of the virus, acting to

CC cleave its replicative proteins from the polypeptide produced from the

CC HCV genome. Inhibitors of the two proteins should be effective as

CC antiviral treatments of HCV infection. This is useful as HCV can lead to

CC chronic liver disease such as cirrhosis, liver failure and liver cancer.

CC The present invention concerns a number of NS3 mutants and NS3-NS4A

CC fusion proteins which can be used to identify inhibitors of this type, as

CC well as enabling structural studies of the protease and

CC protease-inhibitor complexes. This sequence contains the alpha-helix0-1

CC variant.

XX

XX Sequence 197 AA;

SQ

Query Match 93.1%; Score 951; DB 21; Length 197;

Best Local Similarity 94.9%; Pred. No. 5.6e-91;

Matches 187; Conservative 1; Mismatches 7; Indels 2; Gaps 1;

QY 1 MKKGSVVIVGRVING--AYAQTRGEGGCOETSGTRDKNOVEGEVQIVSTAAQTFLA 58

DB 1 MKKGSVVIVGRVINGSLGDTAYAQTRGEGGCOETSGTRDKNOVEGEVQIVSTAAQTFLA 60

QY 59 TCINGVCVTVYHGACTRTIASPKGPVQIOMYTNVDKDLVGNPAPQGSRSLSLTPTCTGSSDLY 118

DB 61 TSINGVLTVYHGACTRTIASPKGPVQIOMYTNVDKDLVGNPAPQGSRSLSLTPTCTGSSDLY 120

QY 119 LVTRHADVIPVRRGDSRGLSPRISYLGSSGGPCLCPAGHAGVIFRAAVCTRGVAK 178

DB 121 LVTRHADVIPVRRGDSRGLSPRISYLGSSGGPCLCPAGHAGVIFRAAVCTRGVAK 180

QY 179 AVDFIPVESLETTMRSP 195

DB 181 AVDFIPVESLETTMRSP 197

RESULT 7

AAB15224

ID AAB15224 standard; protein; 197 AA.

XX

AC AAB15224;

XX

DT 19-DEC-2000 (first entry)

XX

DE Hepatitis C virus NS4A-NS3 fusion protease #6.

XX

XX Hepatitis; NS3 protease; viral replication; chronic liver disease;

KW liver failure; liver cancer; mutant; mutein.

XX

OS Hepatitis C virus.

OS Synthetic.

XX

XX WO2000040707-A1.

PN

XX

PD 13-JUL-2000.

XX

XX 06-JAN-2000; 2000WO-US00345.

PF

XX

PR 08-JAN-1999; 99US-0115271.

XX

PA (BRIM) BRISTOL-MYERS SQUIBB CO.

XX

XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;

PI

```

XX WPI: 2000-465976/40.
DR N-PSDB; AAA73333.
XX
PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
PT amino acid, useful for screening inhibitors that may treat hepatitis C
PT
XX
XX Claim 23; Fig 16; 66pp; English.
XX
CC The present sequence is a mutated version of a fusion protein created
CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
CC proteins are both essential for the replication of the virus, acting to
CC cleave its replicative proteins from the polyprotein produced from the
CC HCV genome. Inhibitors of the two proteins should be effective as
CC antiviral treatments of HCV infection. This is useful, as HCV can lead to
CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
CC The present invention concerns a number of NS3 mutants and NS3-NS4A
CC fusion proteins which can be used to identify inhibitors of this type, as
CC well as enabling structural studies of the protease and
CC protease-inhibitor complexes. This sequence contains the alpha-helix0-7
CC variant.
XX
SQ Sequence 197 AA:
Query Match 92.0%; Score 939; DB 21; Length 197;
Best Local Similarity 93.4%; Pred. No. 1e-89;
Matches 184; Conservative 3; Mismatches 8; Indels 2; Gaps 1;
QY 1 MKKGSVVIVGRIVLNG--AYAQOTRGEQCOETSGTRDKNQVEGEVQIVSTAAQTFLA 58
DB 1 MKKGSVVIVGRINLSGDTAYAQOTRGEQCOETSGTRDKNQVEGEVQIVSTAAQTFLA 60
QY 59 TCINGVCTVYHAGCTRTIASPKGPVIOYNTVDKDLVGPAPQGSRSLLTPCTCGSSDLY 118
DB 61 TSINGVLTWYHAGCTRTIASPKGPVIOYNTVDKDLVGPAPQGSRSLLTPCTCGSSDLY 120
QY 119 LVTRHADVIPVRRGDSRGSLLSPRPISYLGSSGGPLLCPAGHAVGIFRAAVCTRGVAK 178
DB 121 LVTRHADVIPVRRGDSRGSLLSPRPISYLGSSGGPLLCPAGHAVGIFRAAVCTRGVAK 180
QY 179 AVDFIPVESLETTMRSP 195
DB 181 AVDFIPVESLETTMRSP 197
RESULT 8
AAB15225
ID AAB15225 standard; protein; 197 AA.
XX
AC AAB15225;
XX
DT 19-DEC-2000 (first entry)
XX
DE Hepatitis C virus NS4A-NS3 fusion protease #7.
XX
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW liver failure; liver cancer; mutant; mutain.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
XX WO2000040707-A1.
XX
XX 13-JUL-2000.
XX
XX 06-JAN-2000; 2000WO-US00345.
XX
XX 08-JAN-1999; 99US-0115271.
XX
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX

```

```

PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX
XX WPI: 2000-465976/40.
DR N-PSDB; AAA73334.
XX
XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
XX substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
XX amino acid, useful for screening inhibitors that may treat hepatitis C
XX
XX
XX Claim 23; Fig 17; 66pp; English.
XX
XX The present sequence is a mutated version of a fusion protein created
XX using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
XX proteins are both essential for the replication of the virus, acting to
XX cleave its replicative proteins from the polyprotein produced from the
XX HCV genome. Inhibitors of the two proteins should be effective as
XX antiviral treatments of HCV infection. This is useful, as HCV can lead to
XX chronic liver disease such as cirrhosis, liver failure and liver cancer.
XX The present invention concerns a number of NS3 mutants and NS3-NS4A
XX fusion proteins which can be used to identify inhibitors of this type, as
XX well as enabling structural studies of the protease and
XX protease-inhibitor complexes. This sequence contains the alpha-helix0-7
XX variant.
XX
SQ Sequence 197 AA:
Query Match 91.0%; Score 929; DB 21; Length 197;
Best Local Similarity 92.9%; Pred. No. 1.1e-88;
Matches 183; Conservative 3; Mismatches 9; Indels 2; Gaps 1;
QY 1 MKKGSVVIVGRIVLNG--AYAQOTRGEQCOETSGTRDKNQVEGEVQIVSTAAQTFLA 58
DB 1 MKKGSVVIVGRINLSGDTAYAQOTRGEQCOETSGTRDKNQVEGEVQIVSTAAQTFLA 60
QY 59 TCINGVCTVYHAGCTRTIASPKGPVIOYNTVDKDLVGPAPQGSRSLLTPCTCGSSDLY 118
DB 61 TSINGVLTWYHAGCTRTIASPKGPVIOYNTVDKDLVGPAPQGSRSLLTPCTCGSSDLY 120
QY 119 LVTRHADVIPVRRGDSRGSLLSPRPISYLGSSGGPLLCPAGHAVGIFRAAVCTRGVAK 178
DB 121 LVTRHADVIPVRRGDSRGSLLSPRPISYLGSSGGPLLCPAGHAVGIFRAAVCTRGVAK 180
QY 179 AVDFIPVESLETTMRSP 195
DB 181 AVDFIPVESLETTMRSP 197
RESULT 9
AAY44728
ID AAY44728 standard; protein; 191 AA.
XX
AC AAY44728;
XX
DT 04-MAY-2000 (first entry)
XX
DE Hepatitis C virus NS4A-NS3 catalytic domain fusion protein-1.
XX
KW NS3 catalytic domain; NS4A peptide; NS4A-NS3 fusion construct; diagnosis;
KW serine protease; trypsin family; screening; anti-viral compound;
KW treatment; inhibitor; therapeutic.
XX
XX Hepatitis C virus.
XX
XX Key Location/Qualifiers
XX Peptide 1..12
XX /label= NS4A_peptide_1
XX /note= "Covalently attached to amino terminus of NS3
XX catalytic domain"
XX Misc-difference 13
XX /note= "Wild type Proline is replaced with Lysine"
XX
XX WO200001718-A2.

```

```

XX PD 13-JAN-2000.
XX PF 02-JUL-1999; 99WO-US15035.
XX PR 02-JUL-1998; 98US-0091675.
XX PA (UYFL ) UNIV FLORIDA.
XX PI Dunn BM, Bukhtiyarova M;
XX DR WPI; 2000-182103/16.
XX PT Novel polypeptide comprising hepatitis C virus NS4A and NS3 domains,
XX PT useful for screening for compounds useful for the diagnosis and
XX PT treatment of hepatitis C virus.
XX PS Claim 4; Fig 2; 30pp; English.
XX CC The present protein sequence is the fusion polypeptide, comprising the
XX CC hepatitis C virus NS4A peptidic-1 fragment, covalently attached to the
XX CC amino terminus of NS3 catalytic domain. This fusion polypeptide contains
XX CC the NS3 domain expressed in a stable, soluble form. This facilitates the
XX CC use of the polypeptide in direct screening of potential anti-viral
XX CC compounds, that are used for diagnosis and treatment of hepatitis C virus
XX CC infection. It is also used to screen for inhibitors of serine protease
XX CC activity. The polynucleotides are also useful to identify diagnostic or
XX CC therapeutic compounds and for recombinant production of the fusion
XX CC polypeptide.
XX SQ Sequence 191 AA:
Query Match 88.1%; Score 899.5; DB 21; Length 191;
Best Local Similarity 92.1%; Pred. No. 1.3e-85;
Matches 175; Conservative 6; Mismatches 8; Indels 1; Gaps 1;
QY 6 SVIVGVILNGAYAAQTRGEGCOETSGTRKQKQVEGEVOIVSTAAOTFLATCINGVC 65
DB 2 SVIVGVILVUS-KYAQTRGLGCIITSLTGRDRKQVEGEVOIVSTAAOTFLATCINGVR 60
QY 66 WTVYHGAGTRTIA SPKGPVIQMTNVNDKDLVGPAPQGSRLTPCTCGSSDLYLVTRHAD 125
DB 61 WTVYHGAGTRTIA SPKGPVIQMTNVNDKDLVGPAPQGSRLTPCTCGSSDLYLVTRHAD 120
QY 126 VIPVRRRGDSRGSLLSPRPISYIKGSSGGLPCPAGHANGIFRAAVCTRGVAKAVDFIPV 185
DB 121 VIPVRRRGDSRGSLLSPRPISYIKGSSGGLPCPAGHANGIFRAAVCTRGVAKAVDFIPV 180
QY 186 ESLETTMRSP 195
DB 181 ENLETTMRSP 190
QY 186 ESLETTMRSP 195
DB 181 ENLETTMRSP 190
RESULT 10
AAR34009
ID AAR34009 standard; Protein: 2816 AA.
XX AAR34009;
AC AAR34009;
XX 25-MAR-2003 (updated)
DT 26-JUL-1993 (first entry)
XX HCV-1 polypeptide.
XX Polymerase chain reaction; PCR; amplify; primer: hepatitis C virus;
KW HCV; asymptomatic; chronically infected; epitope; viral isolate;
KW domain; immunological; cross-reactive; envelope protein; vaccine;
KW gp53(BVDV)/gp55; hog cholera virus; pestivirus; NS1; flavivirus.
XX OS Hepatitis C virus.
XX PN W09306126-A1.
XX

```

```

PD 01-APR-1993.
XX 11-SEP-1992; 92WO-US07683.
XX 13-SEP-1991; 91US-0759575.
XX (CHIR ) CHIRON CORP.
XX Houghton M, Weiner AJ;
XX WPI; 1993-117468/14.
XX Immuno-reactive hepatitis C virus polypeptide compans. - contg.
XX at least 2 sequences from the first variable domain of distinct
XX HCV isolates
XX PS Disclosure; Fig 9; 106pp; English.
XX CC This sequence represents the entire hepatitis C virus polypeptide.
XX CC HCV is a member of the flavivirus family and appears to encode a basic
XX CC polypeptide domain ("C") at the N-terminal of the viral polypeptide,
XX CC followed by two glycoprotein domains ("E1", "E2/NS1"), upstream of the
XX CC nonstructural genes NS2 through NS5. See also AAQ39134-48, AAR33982-
XX CC 4098 and AAR38088-89.
XX CC (Updated on 25-MAR-2003 to correct PN field.)
XX SQ Sequence 2816 AA:
Query Match 88.0%; Score 898.5; DB 14; Length 2816;
Best Local Similarity 85.8%; Pred. No. 6.1e-84;
Matches 175; Conservative 9; Mismatches 9; Indels 11; Gaps 2;
QY 3 KKGSVIVGV---RIVLNG-----AYAQTRGEGCOETSGTRKQKQVEGEVOIVST 51
DB 1005 RRGEILLGPADGMVMKGRLLAPITAYAQTRGLGCIITSLTGRDRKQVEGEVOIVST 1064
QY 52 AAQTFLATCINGVCWTVYHGAGTRTIA SPKGPVIQMTNVNDKDLVGPAPQGSRLTPCT 111
DB 1065 AAQTFLATCINGVCWTVYHGAGTRTIA SPKGPVIQMTNVNDKDLVGPAPQGSRLTPCT 1124
QY 112 CGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPISYIKGSSGGLPCPAGHANGIFRAAV 171
DB 1125 CGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPISYIKGSSGGLPCPAGHANGIFRAAV 1184
QY 172 CTGVAKAVDFIPVESLETTMRSP 195
DB 1185 CTGVAKAVDFIPVENLETTMRSP 1208
RESULT 11
AAR40120
ID AAR40120 standard; Protein: 3011 AA.
XX AAR40120;
AC AAR40120;
XX 25-MAR-2003 (updated)
DT 27-JAN-1994 (first entry)
XX HCV genomic amino acid sequence isolated from infected human LG.
XX Hepatitis C virus.
XX W09315193-A1.
XX 05-AUG-1993.
XX 29-JAN-1993; 93WO-US00907.
XX 31-JAN-1992; 92US-0830024.

```

XX (ABBO) ABBOTT LAB.
 PA Bode SL, Cascoy JM, Desai SM, Dovare SG, Fraill DE;
 PI Yamaguchi J, Zeck BJ;
 XX WPI; 1993-258673/32.
 DR
 XX
 XX
 PT New plasmid pHCV-162 is a mammalian expression systems for HCV
 PT proteins - useful for diagnosing HCV infection and as vaccines
 PT for preventing HCV infection
 XX
 XX
 PS Example 1: Page 39-49; 100pp; English.
 CC RNA was isolated from the plasma of a HCV seropositive human
 CC (designated "LG") and cDNA was prepared from it. The cDNA was
 CC PCR amplified using specific primers with sequences based
 CC on the prototype HCV-1 cDNA sequence (GENBANK M62321). Further
 CC amplification using nested primers resulted in 7 adjacent HCV DNA
 CC fragments which could be assembled into a full-length sequence. The
 CC DNA sequence was determined and translated into the genomic amino
 CC acid sequence. Comparison of the LG genomic amino acid sequence
 CC with that from HCV-1 showed 134 amino acid differences.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 3011 AA;

Query Match 87.7%; Score 895.5; DB 14; Length 3011;
 Best Local Similarity 85.8%; Pred. No. 1.4e-83;
 Matches 175; Conservative 8; Mismatches 10; Indels 11; Gaps 2;
 OY 3 KGSVVIVG---RIVLNG-----AYAQTRGEGCQETSGTRKKNQVEGEVQIVST 51
 DB 1005 RRGREILGPADGMVSKGWRLLAPITAYAQTRGILGCLITSLGRKKNQVEGEVQIVST 1054
 OY 52 AAQTFLATCINGCVTVYHGAGTRTITASPKGPVIQMTYNDKDLVGPAPQGSRLTPCT 111
 DB 1065 AAQTFLATCINGCVTVYHGAGTRTITASPKGPVIQMTYNDKDLVGPAPQGSRLTPCT 1124
 OY 112 CGSSDLYLVTRHADVIPVRRGDSRGSLLSPRPISYLKSGSGGPLLCPAGHAGVIFRAAV 171
 DB 1125 CGSSDLYLVTRHADVIPVRRGDSRGSLLSPRPISYLKSGSGGPLLCPAGHAGVIFRAAV 1184
 OY 172 CTRGVAKAVDFIPVESLEITMRSP 195
 DB 1185 CTRGVAKAVDFIPVESLEITMRSP 1208

RESULT 12
 AAP92041
 ID AAP92041 standard; protein; 1766 AA.
 XX
 AC AAP92041;

XX 25-MAR-2003 (updated)
 DT 02-MAR-1990 (first entry)

DE Sequence encoded in the hepatitis C virus (HCV) cDNA inserts in clones
 DE 14i, 11b, 7f, 7e, 8h, 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c, 14c, 8f,
 DE 33f, 33g and 39c.

XX Hepatitis C virus (HCV); non-A, non-B hepatitis (HABH)

XX Hepatitis C virus.

XX EP318216-A.

XX 31-MAY-1989.

XX 18-NOV-1988; 88EP-0310922.

XX 18-NOV-1987; 87US-0122714.

PR 30-DEC-1987; 87US-0139886.

PR 26-FEB-1988; 88US-0161072.
 PR 06-MAY-1988; 88US-0191263.
 PR 26-OCT-1988; 88US-0263584.
 PR 14-NOV-1988; 88US-0271450.
 XX
 XX (CHIR) CHIRON CORP.
 PA
 XX
 XX Houghton M, Choo QL, Kuo G;
 XX WPI; 1989-159274/22.
 DR N-PSDB; AAN92097.

XX Purified hepatitis C virus

PT - and associated nucleic acids and polypeptide(s)

XX Claim 13; Figure 26-1, 26-2, 26-3, 26-4, 26-5, 26-6; 139pp; English.

XX It is the sequence encoded in the open reading frame of hepatitis C virus
 CC cDNA inserts in clones 14i, m 11b, 7f, 7e, 8h, 33c, 40b, 37b, 35, 36,
 CC 81, 32, 33b, 25c, 14c, 8f, a33f, 33g and 39c. It is antigenic and could
 CC be used in immunosay reagents and vaccines and to generate antibodies
 CC useful in diagnosis and passive immunotherapy for HCV infection/non-A,
 CC non-B hepatitis.

CC (Updated on 25-MAR-2003 to correct PR field.)

CC (Updated on 25-MAR-2003 to correct PI field.)

XX SQ Sequence 1766 AA;

Query Match 87.6%; Score 894.5; DB 10; Length 1766;
 Best Local Similarity 85.8%; Pred. No. 8.6e-84;
 Matches 175; Conservative 8; Mismatches 10; Indels 11; Gaps 2;

OY 3 KGSVVIVG---RIVLNG-----AYAQTRGEGCQETSGTRKKNQVEGEVQIVST 51
 DB 289 RRGREILGPADGMVSKGWRLLAPITAYAQTRGILGCLITSLGRKKNQVEGEVQIVST 348

OY 52 AAQTFLATCINGCVTVYHGAGTRTITASPKGPVIQMTYNDKDLVGPAPQGSRLTPCT 111
 DB 349 AAQTFLATCINGCVTVYHGAGTRTITASPKGPVIQMTYNDKDLVGPAPQGSRLTPCT 408

OY 112 CGSSDLYLVTRHADVIPVRRGDSRGSLLSPRPISYLKSGSGGPLLCPAGHAGVIFRAAV 171
 DB 409 CGSSDLYLVTRHADVIPVRRGDSRGSLLSPRPISYLKSGSGGPLLCPAGHAGVIFRAAV 468

OY 172 CTRGVAKAVDFIPVESLEITMRSP 195

DB 469 CTRGVAKAVDFIPVENLEITMRSP 492

RESULT 13

AAP90158

ID AAP90158 standard; protein; 1786 AA.

XX AAP90158;

XX 25-MAR-2003 (updated)

DT 10-NOV-1989 (first entry)

XX Protein sequence of hepatitis c virus composite cDNA.

XX Hepatitis C virus; vaccine.

XX Pan troglodytes.

XX GB2212511-A.

XX 26-JUL-1989.

XX 18-NOV-1988; 88GB-0027024.

XX 18-NOV-1987; 87US-0122714.

PR 30-DEC-1987; 87US-0139886.

PR 26-FEB-1988; 88US-0161072.

```

PR 26-OCT-1988; 88US-0263584.
XX (CHIR ) CHIRON CORPORATION.
PA Houghton M, Choo QL, Kuo G;
PI WPI; 1989-215054/30.
XX N-PSDB; AAN90327.
XX Hepatitis C virus gene - used for prodn. of polynucleotide probes,
PT polypeptide(s) and antibodies for diagnosis, prevention and treatment
PT of infection.
XX Disclosure: fig 26; 30pp; English.
XX The sequence is encoded by the composite cDNA of AAN90327. These
CC antigens react with antibodies in patients with non-B hepatitis
CC (NANBH). They can be used to diagnose HCV-induced NANBH, to raise
CC antibodies for immunoassay or treatment, or to produce vaccines.
CC (Updated on 25-MAR-2003 to correct PR field.)
XX Sequence 1786 AA;
XX
Query Match 87.68; Score 894.5; DB 10; Length 1786;
Best Local Similarity 85.8%; Pred. No. 8.7e-84;
Matches 175; Conservative 8; Mismatches 10; Indels 11; Gaps 2;
QY 3 KKGSVWIVG---RIVLNG-----AYAOQTRGEGCCOETSGTGRDKNOVEGEVIVST 51
DB 289 RRGREILLGPGADGVMVSGWRLAPITAYAOQTGRLGCIITSLTGRDKNOVEGEVIVST 348
QY 52 AAQTFLATCINGVCWTVYHGAGTGTIASPKGPVIQMYTNVDKDLVGPAPQGSRSITPCT 111
DB 349 AAQTFLATCINGVCWTVYHGAGTGTIASPKGPVIQMYTNVDQDLVGPAPQGSRSITPCT 408
QY 112 CGSSDLYLVTRHADVIPVRRGDSRGLSPRPISYLVKSSGGPILCPAGHAGVIFRAAV 171
DB 409 CGSSDLYLVTRHADVIPVRRGDSRGLSPRPISYLVKSSGGPILCPAGHAGVIFRAAV 468
QY 172 CTRGVAKAVDFIPVESLETTMRSP 195
DB 469 CTRGVAKAVDFIPVENLETTMRSP 492
RESULT 14
AAP90164
ID AAP90164 standard; protein: 2261 AA.
XX AAP90164;
XX 25-MAR-2003 (updated)
DT 01-NOV-1989 (first entry)
XX Peptide encoded by composite hepatitis C virus cDNA.
DE Hepatitis C virus; clone 12f; clone 15e; probe: vaccine.
XX Pan troglodytes.
XX GB2212511-A.
XX 26-JUL-1989.
XX 18-NOV-1988; 88GB-0027024.
XX 18-NOV-1987; 87US-0122714.
PR 30-DEC-1987; 87US-0139886.
PR 26-FEB-1988; 88US-0161072.
PR 18-NOV-1987; 87US-0122714.
PR 30-DEC-1987; 87US-0139886.
PR 26-FEB-1988; 88US-0161072.
PR 26-OCT-1988; 88US-0263584.
XX (CHIR ) CHIRON CORPORATION.
PA Houghton M, Choo QL, Kuo G;
PI WPI; 1989-215054/30.
XX N-PSDB; AAN90331.
XX Hepatitis C virus gene - used for prodn. of polynucleotide probes,
PT polypeptide(s) and antibodies for diagnosis, prevention and
PT treatment of infection.
XX Disclosure: fig 32; 235pp; English.
XX The sequence is the peptide encoded by the composite hepatitis C
CC virus (HCV) cDNA of AAN90331. The polypeptides are used to diagnose
CC HCV-induced NANBH, to raise antibodies for immunoassay or treatment,
CC or to produce vaccines.
CC (Updated on 25-MAR-2003 to correct PR field.)
XX Sequence 2261 AA;
XX
Query Match 87.6%; Score 894.5; DB 10; Length 2261;
Best Local Similarity 85.8%; Pred. No. 1.2e-83;
Matches 175; Conservative 8; Mismatches 10; Indels 11; Gaps 2;
QY 3 KKGSVWIVG---RIVLNG-----AYAOQTRGEGCCOETSGTGRDKNOVEGEVIVST 51
DB 380 RRGREILLGPGADGVMVSGWRLAPITAYAOQTGRLGCIITSLTGRDKNOVEGEVIVST 439
QY 52 AAQTFLATCINGVCWTVYHGAGTGTIASPKGPVIQMYTNVDKDLVGPAPQGSRSITPCT 111
DB 440 AAQTFLATCINGVCWTVYHGAGTGTIASPKGPVIQMYTNVDQDLVGPAPQGSRSITPCT 499
QY 112 CGSSDLYLVTRHADVIPVRRGDSRGLSPRPISYLVKSSGGPILCPAGHAGVIFRAAV 171
DB 500 CGSSDLYLVTRHADVIPVRRGDSRGLSPRPISYLVKSSGGPILCPAGHAGVIFRAAV 559
QY 172 CTRGVAKAVDFIPVESLETTMRSP 195
DB 560 CTRGVAKAVDFIPVENLETTMRSP 583

```

```

RESULT 15
AAP92047
ID AAP92047 standard; protein: 2301 AA.
XX AAP92047;
XX 25-MAR-2003 (updated)
DT 02-MAR-1990 (first entry)
XX Sequence encoded in the hepatitis C virus (HCV) cDNA inserts in clones
DE 12f through 15e.
XX Hepatitis C virus (HCV); non-A, non-B hepatitis (NANBH).
XX Hepatitis C virus.
XX EP318216-A.
XX 31-MAY-1989.
XX 18-NOV-1988; 88EP-0310922.
XX 18-NOV-1987; 87US-0122714.
PR 30-DEC-1987; 87US-0139886.
PR 26-FEB-1988; 88US-0161072.
PR 06-MAY-1988; 88US-0191263.
PR 26-OCT-1988; 88US-0263584.
PR 14-NOV-1988; 88US-0271450.
XX (CHIR ) CHIRON CORP.
XX Houghton M, Choo QL, Kuo G;
XX WPI; 1989-159274/22.

```

DR N-PSDB; AAN92103.
XX
PT Purified hepatitis C virus
PT - and associated nucleic acids and polypeptide(s)
XX
PS Claim 13; Figure 32-1 - 32-7; 139 pp; English.
XX
CC It is the sequence encoded in the open reading frame of hepatitis C virus
CC (HCV) cDNA inserts in clones 12f through 15e. It is antigenic and could
CC be used in immunosay reagents and vaccines and to generate antibodies
CC useful in diagnosis and passive immunotherapy for HCV infection/non-A,
CC non-B hepatitis.
CC (Updated on 25-MAR-2003 to correct PR field.)
CC (Updated on 25-MAR-2003 to correct PI field.)
XX
SQ Sequence 2301 AA:

Query Match 87.6%; Score 894.5; DB 10; Length 2301;
Best Local Similarity 85.8%; Pred. No. 1.2e-83;
Matches 175; Conservative 8; Mismatches 10; Indels 11; Gaps 2;

QY 3 KKGSVVIVG---RIVLNG-----AVAQOTRGECCQFTSOTCRDNKNOVEGFVIVST 51
DB 380 RRGREILGPADGMVSKGWRLAPITAYAQOTRGLLGCITISLGRDNQVEGEVQIVST 439

QY 52 AAQTFLATCINGVCWTVYHGAGIRTIASPKGPIVQIOMYTNVDKDLVGPAPQGSRSLEPCT 111
DB 440 AAQTFLATCINGVCWTVYHGAGIRTIASPKGPIVQIOMYTNVDQDLVGPAPQGSRSLEPCT 499

QY 112 CGSSDLYLVTRHADVIPVRRGDSRGLSPRPISYLGSSGGPLLCPAGHAVGIFRAAV 171
DB 500 CGSSDLYLVTRHADVIPVRRGDSRGLSPRPISYLGSSGGPLLCPAGHAVGIFRAAV 559

QY 172 CTRGVAKAVDFIPVESLETMRSP 195
DB 560 CTRGVAKAVDFIPVENLETMRSP 583

Search completed: August 30, 2003, 19:12:22
Job time : 47.1697 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 19:02:22 ; Search time 16.0488 Seconds
(without alignments)
1168.492 Million cell updates/sec

Title: US-09-965-594-12

Perfect score: 1021

Sequence: 1 MKKKGSVVIGRVLNGAYA.....VAKAVDFIPVESLETIMKSP 195

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_76: *
1: PIR1: *
2: PIR2: *
3: PIR3: *
4: PIR4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Length	DR	ID	Description
1	894.5	87.6	3011	1	GNMVC3	genome polyprotein
2	893.5	87.5	3011	1	S40770	genome polyprotein
3	879.5	86.1	3011	1	GNMVC4	genome polyprotein
4	847	83.0	3010	1	GNMVTW	genome polyprotein
5	842	82.5	3010	1	GNMVTG	genome polyprotein
6	838	82.1	3010	1	GNMVCJ	genome polyprotein
7	837	82.0	3010	1	A45573	genome polyprotein
8	821	80.4	3010	1	SI8030	genome polyprotein
9	761	74.5	3014	1	JC5620	genome polyprotein
10	670.5	65.7	3033	1	JQ1303	genome polyprotein
11	668.5	65.5	3033	1	GNMVCJ	genome polyprotein
12	260	25.5	2970	2	T08839	polyprotein - marm
13	258	25.3	3005	2	T08841	polyprotein - dour
14	85.5	8.4	495	2	B71360	hypothetical prote
15	85.5	8.4	1334	2	AB1775	hypothetical prote
16	82	8.0	452	2	I39383	angio-associated m
17	80.5	7.9	590	2	B81104	nitrate/nitrite se
18	80.5	7.9	590	2	C81911	nitrate/nitrite se
19	79.5	7.8	642	1	VCMVFG	env polyprotein -
20	79	7.7	259	1	IOH01	insulin-like growt
21	79	7.7	354	2	T49806	hypothetical prote
22	79	7.7	404	2	A46165	envelope surface g
23	78.5	7.7	209	2	H83144	probable aromatic
24	78.5	7.7	981	2	T18234	beta transducin ho
25	77.5	7.6	398	2	B71284	probable periplasm
26	77.5	7.6	2663	1	S28261	centromere protein
27	77.5	7.6	3507	2	T34513	hypothetical prote
28	76.5	7.5	270	2	T06118	hypothetical prote
29	76.5	7.5	393	2	E95261	serine proteinase

30 76.5 7.5 397 2 B98127 serine proteinase
31 76 7.4 140 2 C72705 hypothetical prote
32 76 7.4 574 2 A84782 hypothetical prote
33 75.5 7.4 859 2 T35785 probable beta-gluc
34 75.5 7.4 882 2 S41034 hypothetical prote
35 75 7.3 639 1 VCMVSA env polyprotein pr
36 75 7.3 1293 2 T30871 orsellinic acid sy
37 74.5 7.3 415 2 S70401 zona pellucida gly
38 74.5 7.3 492 2 AH1030 probable exported
39 74.5 7.3 755 2 S23441 hypothetical prote
40 74.5 7.3 1176 2 T18042 ice nucleation pro
41 74 7.2 239 2 H89966 serine proteinase
42 74 7.2 239 2 G87265 conserved hypothet
43 74 7.2 603 1 VCFVER env polyprotein -
44 73.5 7.2 317 2 S76618 hypothetical prote
45 73.5 7.2 566 2 H84203 phosphate ABC tran

ALIGNMENTS

genome polyprotein - hepatitis C virus (strain HCV-1)
N:Contains: capsid protein C; envelope protein M; hepatitisin (EC 3.4.21.98) (nonstr
Protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 30-Sep-1992 #sequence:revision 30-Sep-1992 #text_change 19-Jan-2001
C:Accession: A39166; PQ0403; PQ0404
R:Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.;
Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991
A:Title: Genetic organization and diversity of the hepatitis C virus.
A:Reference number: A39166; MUID:91172826; PMID:1848704
A:Accession: A39166
A:Molecule type: mRNA
A:Residues: 1-3011 <CHO>
A:Cross-references: GB:M62321; NID:g329873; PIDN:AAA45676.1; PID:g329874
R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peuchner, J.F.; Follett, E.; Yap,
J. Gen. Virol. 73, 1131-1141, 1992
A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship
A:Reference number: PQ0393; MUID:92268871; PMID:1316939
A:Accession: PQ0403
A:Molecule type: genomic RNA
A:Residues: 1577-1633 <CHA>
A:Cross-references: DDBJ:D10128
A:Experimental source: isolates E-b16
A:Accession: PQ0404
A:Status: preliminary
A:Molecule type: genomic RNA
A:Residues: 1577-1633 <CH2>
A:Experimental source: isolates E-b17
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstru
F:1-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitisin #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,20

Query Match 87.6% Score 894.5 DB 1; Length 3011;

Best Local Similarity 85.8% Pred. No. 1.7e-74;

Matches 175; Conservative 8; Mismatches 10; Indels 11; Gaps 2;

QY 3 KKGSVVIG---RIVLNG-----AYAQTRGECQETSOTGRKKNQVEGVQIVST 51
::: ::::: ::|
||||||| ||| |||||||


```
Db 1005 RRCKEILLGPADQWVGKQWLLAPITAYAAQQTGRLGLGCIITSLTGDKNQVEGVQIVST 1064
Qy 52 AAQTFLATCINGVCWTVYHGAGTRTIASPKGPVIQMYINWCKDLVGVWPAQGSRLTPTCT 111
Db 1065 AAQTFLATCINGVCWTVYHGAGTRTIASPKGPVIQMYINWCKDLVGVWPAQGSRLTPTCT 1124
Qy 112 CGSSDLVLTVRHADVIPIVRRRGDSRGSLSPRPISYLKSSGGPGLCPAGHAGVIFRAAV 171
Db 1125 CGSSDLVLTVRHADVIPIVRRRGDSRGSLSPRPISYLKSSGGPGLCPAGHAGVIFRAAV 1184
Qy 172 CTRGVAKAVDFIPVESLETTMRSP 195
Db 1185 CTRGVAKAVDFIPVESLETTMRSP 1208

RESULT 2
genome polyprotein - hepatitis C virus
N:Contains: capsid protein C; envelope protein M; hepatitis M; hepatitis C virus (strain H)
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C:Accession: A36814; A41546
R:Inchauspe, G.; Zebedes, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
submitted to GenBank, July 1992
A:Description: Genomic structure of the human prototype strain H of hepatitis C virus
A:Reference number: A41546; MUID:92052256; PMID:1658600
A:Contents: annotation
A:Note: neither amino acid nor nucleotide sequence is given
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruct
F:1-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEP>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis M #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240

submitted to the EMBL Data Library, March 1992
A:Reference number: S40770
A:Accession: S40770
A:Molecule type: genomic RNA
A:Residues: 1-3011 <KOK>
A:Cross-references: EMBL:D10749; NID:g221586; PIDN:BAA01582.1; PID:g221587
R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda,
Jpn. J. Exp. Med. 60, 167-177, 1990
A:Title: The 5'-terminal sequence of the hepatitis C virus genome.
A:Reference number: PC1284; MUID:91013116; PMID:2170712
A:Accession: PC1285
A:Molecule type: genomic RNA
A:Residues: 1-513 <OK2>
A:Cross-references: GB:D00831; NID:g221511; PIDN:BAA00705.1; PID:g221512
A:Experimental source: isolate HC-J1
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEP>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis M #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 87.5%; Score 893.5; DB 1: Length 3011;
Best Local Similarity 85.8%; Pred. No. 2,1e-74;
Matches 175; Conservative 7; Mismatches 11; Indels 11; Gaps 2;

Qy 3 KKGSVTVG---RIVLNG-----AYAOOTRGECCQETISOTGRDKNQVEGVQIVST 51
Db 1005 RKGREIILGPADGVMVSKGWLLAPITAYAAOQTGRLGCIITSLTGDKNQVEGVQIVST 1064
Qy 52 AAQTFLATCINGVCWTVYHGAGTRTIASPKGPVIQMYINWCKDLVGVWPAQGSRLTPTCT 111
Db 1065 AAQTFLATCINGVCWTVYHGAGTRTIASPKGPVIQMYINWCKDLVGVWPAQGSRLTPTCT 1124
Qy 112 CGSSDLVLTVRHADVIPIVRRRGDSRGSLSPRPISYLKSSGGPGLCPAGHAGVIFRAAV 171
Db 1125 CGSSDLVLTVRHADVIPIVRRRGDSRGSLSPRPISYLKSSGGPGLCPAGHAGVIFRAAV 1184
Qy 172 CTRGVAKAVDFIPVESLETTMRSP 195
Db 1185 CTRGVAKAVDFIPVESLETTMRSP 1208
```

```
RESULT 3
GNWVCH
genome polyprotein - hepatitis C virus (strain H)
N:Contains: capsid protein C; envelope protein M; hepatitis M; hepatitis C virus (strain H)
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C:Accession: A36814; A41546
R:Inchauspe, G.; Zebedes, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
submitted to GenBank, July 1992
A:Description: Genomic structure of the human prototype strain H of hepatitis C virus
A:Reference number: A41546; MUID:92052256; PMID:1658600
A:Contents: annotation
A:Note: neither amino acid nor nucleotide sequence is given
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruct
F:1-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEP>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis M #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240

Query Match 86.1%; Score 879.5; DB 1: Length 3011;
Best Local Similarity 83.8%; Pred. No. 4,1e-73;
Matches 171; Conservative 9; Mismatches 13; Indels 11; Gaps 2;

Qy 3 KKGSVTVG---RIVLNG-----AYAOOTRGECCQETISOTGRDKNQVEGVQIVST 51
Db 1005 RRGQEIILGPADGVMVSKGWLLAPITAYAAOQTGRLGCIITSLTGDKNQVEGVQIVST 1064
Qy 52 AAQTFLATCINGVCWTVYHGAGTRTIASPKGPVIQMYINWCKDLVGVWPAQGSRLTPTCT 111
Db 1065 AAQTFLATCINGVCWTVYHGAGTRTIASPKGPVIQMYINWCKDLVGVWPAQGSRLTPTCT 1124
Qy 112 CGSSDLVLTVRHADVIPIVRRRGDSRGSLSPRPISYLKSSGGPGLCPAGHAGVIFRAAV 171
Db 1125 CGSSDLVLTVRHADVIPIVRRRGDSRGSLSPRPISYLKSSGGPGLCPAGHAGVIFRAAV 1184
Qy 172 CTRGVAKAVDFIPVESLETTMRSP 195
Db 1185 CTRGVAKAVDFIPVESLETTMRSP 1208

RESULT 4
GNWVCH
genome polyprotein - hepatitis C virus (strain Taiwan)
N:Contains: capsid protein C; envelope protein M; hepatitis M; hepatitis C virus (strain Taiwan)
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C:Accession: A40244
R:Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.
Virology 186, 102-113, 1992
```

Matches 154; Conservative 14; Mismatches 10; Indels 0; Gaps 0;

Qy 18 AYAOOTRGECCQETSOTGRDNKQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRTI 77
||| ||||| ||| |
Db 1031 AYSOOTRGLGCIITSITGRDNKQVEGEVQIVSTATQSFLATCINGVCWTVYHGAGSKTL 1090
||| ||||| ||| |

Qy 78 ASPKGPVIOMTYNDXLDLVGHAPPOGSRSLPCTCGSSDLYLVTTHADVIPVRRRGDSRG 137
| :|||: |||||: |||||: | :|||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Db 1091 AAPKGPIQMITYNDVDQLVGMPKPFGARSLPCTCGSSDLYLVTTHADVIPVRRRGDSRG 1150
| :|||: |||||: |||||: | :|||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:

Qy 138 SLLSPRPISYLKSGGGPLLCPAGHAVGIFRAAIVCTRGAKAVDFIPVESLETTMRSP 195
||| ||||| ||| |
Db 1151 SLLSPRPVSILKSGGGPLLCFPHAVGIFRAAIVCTRGAKAVDFVPVESMETTMRSP 1208
||| ||||| ||| |

RESULT 6

GNWVCJ

genome polypeptin - hepatitis C virus (strain J)

N:Contains: capsid protein C; envelope protein M; major envelope protein E; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: Hepatitis C virus

C>Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 19-Jan-2001

A:Accession: A39253; PS0086

R:Kato, N.; Hijikata, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shino-oka, K.; Shimotohno, K. Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990

A:Title: Molecular cloning of the human hepatitis C virus genome from Japanese patient

A:Reference number: A39253; PMID:91088550; PMID:2175903

A:Accession: A39253

A:Molecule type: genomic RNA

A:Residues: 1-3010 <RAT>

A:Cross-references: GB:D90208; NID:g221610; PIDN:BAA14233.1; PID:g221611

R:Kato, N.; Ohkoshi, S.; Shimotohno, K. Proc. Jpn. Acad. 65B, 219-223, 1989

A:Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence homology with those of the non-A, non-B hepatitis viral genome

A:Reference number: PS0085

A:Accession: PS0086

A:Molecule type: genomic RNA

A:Residues: 2650-2707 <KA2>

A:Experimental source: Japanese isolate

C:Comment: The cleavage sites of this polypeptide have not been determined.

C:Superfamily: hepatitis C virus genome polypeptide

C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; signal peptide; capsid protein C #status predicted <CPC>

F:2-115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: envelope protein M #status predicted <EPM>

F:192-389/Product: major envelope protein E #status predicted <EEP>

F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F:1007-1615/Product: hepacivirus #status predicted <NS3>

F:1230-1237/Region: nucleotide-binding motif A (P-loop)

F:1312-1317/Region: nucleotide-binding motif B

F:1316-1319/Region: DXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <NA4>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <NB>

F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,234,250,305,325,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2:7

Query Match 82.1%; Score 838; DB 1; Length 3010;
Best Local Similarity 85.4%; Pred No. 3e-69;
Matches 152; Conservative 16; Mismatches 10; Indels 0; Gaps 0;

Qy 18 AYAOOTRGECCQETSOTGRDNKQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRTI 77
||| ||||| ||| |
Db 1031 AYSOOTRGLGCIITSITGRDNKQVDGEVOVLSTATQSFLATCINGVCWTVYHGAGSKTL 1090
||| ||||| ||| |

Qy 78 ASPKGPVIOMTYNDXLDLVGHAPPOGSRSLPCTCGSSDLYLVTTHADVIPVRRRGDSRG 137
| :|||: |||||: |||||: | :|||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Db 1091 AAPKGPIQMITYNDVDQLVGMPKPFGARSLPCTCGSSDLYLVTTHADVIPVRRRGDSRG 1150
| :|||: |||||: |||||: | :|||: |||||: |||||: |||||: |||||: |||||: |||||:

Qy 138 SLLSPRPISYLKSGGGPLLCPAGHAVGIFRAAIVCTRGAKAVDFIPVESLETTMRSP 195
||| ||||| ||| |
Db 1151 SLLSPRPISYLKSGGGPLLCFPHAVGIFRAAIVCTRGAKAVDFIPVESMETTMRSP 1208
||| ||||| ||| |

RESULT 7

A45573
genome polyprotein - hepatitis C virus (strain J7)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-1006/Product: nonstructural protein NS1 #status predicted <NS1>
F:1007-1615/Product: nonstructural protein NS2 #status predicted <NS2>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
Query Match 82.0%; Score 837; DB 1; Length 3010;
Best Local Similarity 86.5%; Pred. No. 3 7e-69;
Matches 154; Conservative 13; Mismatches 11; Indels 0; Gaps 0;
QY 18 AYAAQTRGEGCGEQTSGTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRTI 77
DB 1031 AYAAQTRGLLGCVITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGSKTL 1090
QY 78 ASPKGPVIQMTNVKDLVGVPAQGSRLTPTCGSSDLYLTVRHADVIVPRRGDSRG 137
DB 1091 AGPKGPITQMTNVDDLVGVPAQGSRLTPTCGSSDLYLTVRHADVIVPRRGDSRG 1150
QY 138 SLLSPRPISYLGSGSGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMRSP 195
DB 1151 SLLSPRPVSYLGSGSGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMRSP 1208

RESULT 8

S18030
genome polyprotein - hepatitis C virus (isolate JK1)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; se
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-1006/Product: nonstructural protein NS1 #status predicted <NS1>
F:1007-1615/Product: nonstructural protein NS2 #status predicted <NS2>
F:1231-1238/Region: nucleotide-binding motif A (P-loop)
F:1313-1318/Region: nucleotide-binding motif B
F:1317-1320/Region: DEXH motif
F:1617-1863/Product: nonstructural protein NS4a #status predicted <NS4a>
F:1864-2014/Product: nonstructural protein NS4b #status predicted <NS4b>
Query Match 82.0%; Score 837; DB 1; Length 3010;
Best Local Similarity 86.5%; Pred. No. 3 7e-69;
Matches 154; Conservative 13; Mismatches 11; Indels 0; Gaps 0;
QY 18 AYAAQTRGEGCGEQTSGTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRTI 77
DB 1031 AYAAQTRGLLGCVITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGSKTL 1090
QY 78 ASPKGPVIQMTNVKDLVGVPAQGSRLTPTCGSSDLYLTVRHADVIVPRRGDSRG 137
DB 1091 AGPKGPITQMTNVDDLVGVPAQGSRLTPTCGSSDLYLTVRHADVIVPRRGDSRG 1150
QY 138 SLLSPRPISYLGSGSGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMRSP 195
DB 1151 SLLSPRPVSYLGSGSGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMRSP 1208

A:Accession: S33570
A:Molecule type: genomic RNA
A:Residues: 1-547,'T',549-621,'V',623-624,'S',626-652,'DL',655-761,'T',763-782 <HOW>
A:Cross-references: EMBL:X61591
A:Note: this sequence is inconsistent with the nucleotide translation
A:Note: the authors translated the codon AGG for residue 43 as Pro, TGG for residue 3
as Trp, and TTC for residue 771 as Ser
A:Note: sequence extracted from NCBI backbone (NCBIN:121747, NCBI:121748)
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; se
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-1006/Product: nonstructural protein NS1 #status predicted <NS1>
F:1007-1615/Product: nonstructural protein NS2 #status predicted <NS2>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F:196-209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate
Query Match 80.4%; Score 821; DB 1; Length 3010;
Best Local Similarity 85.4%; Pred. No. 1 1e-67;
Matches 152; Conservative 13; Mismatches 13; Indels 0; Gaps 0;
QY 18 AYAAQTRGEGCGEQTSGTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRTI 77
DB 1031 AYAAQTRGEGCGEQTSGTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGSKTL 1090
QY 78 ASPKGPVIQMTNVKDLVGVPAQGSRLTPTCGSSDLYLTVRHADVIVPRRGDSRG 137
DB 1091 AGPKGPITQMTNVDDLVGVPAQGSRLTPTCGSSDLYLTVRHADVIVPRRGDSRG 1150
QY 138 SLLSPRPISYLGSGSGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMRSP 195
DB 1151 SLLSPRPVSYLGSGSGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMRSP 1208

RESULT 9

JC5620
genome polyprotein - hepatitis C virus (isolate EUH1480)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; se
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-1006/Product: nonstructural protein NS1 #status predicted <NS1>
F:1007-1615/Product: nonstructural protein NS2 #status predicted <NS2>
F:1231-1238/Region: nucleotide-binding motif A (P-loop)
F:1313-1318/Region: nucleotide-binding motif B
F:1317-1320/Region: DEXH motif
F:1617-1863/Product: nonstructural protein NS4a #status predicted <NS4a>
F:1864-2014/Product: nonstructural protein NS4b #status predicted <NS4b>
Query Match 82.0%; Score 837; DB 1; Length 3010;
Best Local Similarity 86.5%; Pred. No. 3 7e-69;
Matches 154; Conservative 13; Mismatches 11; Indels 0; Gaps 0;
QY 18 AYAAQTRGEGCGEQTSGTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRTI 77
DB 1031 AYAAQTRGLLGCVITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGSKTL 1090
QY 78 ASPKGPVIQMTNVKDLVGVPAQGSRLTPTCGSSDLYLTVRHADVIVPRRGDSRG 137
DB 1091 AGPKGPITQMTNVDDLVGVPAQGSRLTPTCGSSDLYLTVRHADVIVPRRGDSRG 1150
QY 138 SLLSPRPISYLGSGSGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMRSP 195
DB 1151 SLLSPRPVSYLGSGSGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMRSP 1208

F:2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>
F:2210-2249/Region: interferon sensitivity determining #status predicted

Query Match 74.5%; Score 761; DB 1; Length 3014;
Best Local Similarity 77.5%; Pred. No. 4.3e-62;
Matches 138; Conservative 19; Mismatches 21; Indels 0; Gaps 0;

QY 18 AYAQTRGEECCQTSOTGRDNKQVEGEVQIVSTAQTFLATCINGVCVTVYHGAGTRTI 77
DB 1032 AYAQTRGVLCAGVILSLTGRDKNEAGEVQVFLSTATQTFGLCINGVWTVLPHGAGSKTL 1091

QY 78 ASPKGPVQMTYNDKDLVGHWPAPQGSRLTPCTCGSSDLYLTVRHADVPVRRRGDSRG 137
DB 1092 AGPKGPVQMTYNDKDLVGHWPSPGKSLTRCTCGSADLYLTVRHADVPVRRRGDTA 1151

QY 138 SLLSPRISYLGSGGPGLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMRSP 195
DB 1152 SLLSPRISYLGSGGPGIMCPSGHVGVFRAAVCTRGVAKAVEFVPEVLETTMRSP 1209

RESULT 10
J01303
genome polyprotein - hepatitis C virus (isolate HC-J6)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 17-Nov-2000
C:Accession: J01303
R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Kurai, K.; Mizuka, H.; Machida, A.; Miyakawa, Y.
J. Gen. Virol. 72, 2697-2704, 1991
A:Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human
A:Reference number: J01303; MUID:92044440; PMID:1658196
A:Accession: J01303
A:Molecule type: genomic RNA
A:Residues: 1-3033 <OKA>
A:Cross-references: GB:D00944; NID:g221650; PIDN:BAA00792.1; PID:g221651
A:Experimental source: isolate HC-J6 from a Japanese individual
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; transmembrane
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-733/Product: nonstructural protein NS1 #status predicted <NS1>
F:1011-1619/Product: hepatitis C virus NS2 #status predicted <NS2>
F:1316-1321/Region: nucleotide-binding motif B
F:1320-1323/Region: DEXH motif
F:1620-1866/Product: nonstructural protein NS4a #status predicted <NS4A>
F:1867-2017/Product: nonstructural protein NS4b #status predicted <NS4B>
F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,305,325,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,28

Query Match 65.7%; Score 670.5; DB 1; Length 3033;
Best Local Similarity 62.1%; Pred. No. 1.1e-53;
Matches 126; Conservative 27; Mismatches 33; Indels 17; Gaps 1;

QY 10 VGRIVLNG-----AYAQTRGEECCQTSOTGRDNKQVEGEVQIVSTA 52
DB 1010 LGREVLGPGADYTSKGSWLLAPITAYAQTRGLLGTIVSVMTGRDTEQAGEIQVLSTV 1069

QY 53 AOTFLATCINGVCVTVYHGAGTRTIASPKGVIOMYTNVDKDLVGPAPQGSRLTPTCTC 112
DB 1070 TQFLGTGTSISGLVTVYHGAGNKTLAGSGPVTOMYSSAEGDLVGPSPPGTKSLDPTCTC 1129

QY 113 GSSDLYLTVRHADVPVRRRGDSRLSPRISYLGSGGPGLLCPAGHAGVIFRAAVC 172
DB 1130 GAVDLYLTVRNADVPVRRRGDKRGLLSPRLSTLKGSSGGPVLCPRGHAGVIFRAAVC 1189

QY 173 TRGVAKAVDFIPVESLETTMRSP 195
DB 1190 SRGVAKSIDFIPVETLDTVTRSP 1212

RESULT 11
GNMWJ8
genome polyprotein - hepatitis C virus (strain HC-J8)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C:Accession: A40250; PQ0397; PQ0559
R:Okamoto, H.; Kurai, K.; Okada, S.I.; Yamamoto, K.; Mizuka, H.; Tanaka, T.; Fukuda, Y.
Virol. 188, 331-341, 1992
A:Title: Full-length sequence of a hepatitis C virus genome having poor homology to
A:Reference number: A40250; MUID:92230232; PMID:1314459
A:Accession: A40250
A:Molecule type: genomic RNA
A:Residues: 1-3033 <OKA>
A:Cross-references: GB:D10988; GB:D01221; NID:g221608; PIDN:BAA01761.1; PID:g221609
R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, J. Gen. Virol. 73, 1131-1141, 1992
A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship
A:Reference number: PQ0393; MUID:92268871; PMID:1316939
A:Accession: PQ0397
A:Molecule type: genomic RNA
A:Residues: 2678-2754 <CHA>
A:Cross-references: DDBJ:D10134
A:Experimental source: isolate E-bl2
R:Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Mori, S.; Hijikata, M.; Shimoto
Biochem. Biophys. Res. Commun. 181, 279-285, 1991
A:Title: Distribution of plural HCV types in Japan.
A:Reference number: PQ0554; MUID:92068204; PMID:1720309
A:Accession: PQ0559
A:Molecule type: mRNA
A:Residues: 2678-2729 <XAT>
A:Cross-references: GB:D10562; GB:D90518; NID:g221523; PIDN:BAA01418.1; PID:g221524
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
F:1-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-733/Product: nonstructural protein NS1 #status predicted <NS1>
F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>
F:1011-1619/Product: hepatitis C virus NS3 #status predicted <NS3>
F:1334-1241/Region: nucleotide-binding motif A (P-loop)
F:1316-1321/Region: nucleotide-binding motif B
F:1320-1323/Region: DEXH motif
F:1620-1866/Product: nonstructural protein NS4a #status predicted <NS4A>
F:1867-2017/Product: nonstructural protein NS4b #status predicted <NS4B>
F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,233,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,20

Query Match 65.5%; Score 668.5; DB 1; Length 3033;
Best Local Similarity 62.6%; Pred. No. 1.8e-53;
Matches 127; Conservative 25; Mismatches 34; Indels 17; Gaps 1;

QY 10 VGRIVLNG-----AYAQTRGEECCQTSOTGRDNKQVEGEVQIVSTA 52
DB 1010 LGREVLGPGADYTSKGSWLLAPITAYTQTRGLGLGAIYVSLTGRDNEAQGVQLSSV 1069

QY 53 AOTFLATCINGVCVTVYHGAGTRTIASPKGVIOMYTNVDKDLVGPAPQGSRLTPTCTC 112
DB 1070 TQFLGTGTSISGLVTVYHGAGNKTLAGSGPVTOMYSSAEGDLVGPSPPGTKSLDPTCTC 1129

QY 113 GSSDLYLTVRHADVPVRRRGDSRLSPRISYLGSGGPGLLCPAGHAGVIFRAAVC 172
DB 1130 GAVDLYLTVRNADVPVRRRGDKRGLLSPRLSTLKGSSGGPVLCPRGHAGVIFRAAVC 1189

QY 173 TRGVAKAVDFIPVESLETTMRSP 195
DB 1490 ARGVAKSIDFIPVESLDVATRP 1212

RESULT 12
T08839
polyprotein - marmoset hepatitis GB virus A

```
Query Match      8.4%; Score 85.5; DB 2; Length 1334;  
Best Local Similarity 27.0%; Pred. No. 11;  
Matches 58; Conservative 19; Mismatches 83; Indels 55; Gaps 10;
```

Qy 3 KKG---SVVIVGRVLNGAYAAQQTRGEGCGETSQTGRDNKNQVEGVQLTSTAATFTPLAT 59
||| || | : || | : || | : || | : || | :

Db	738	KGILQSLKIVDELVSVMGYPQTIIVVEMARENQTTCKGKNNNSRPRYKSLEKAKEFGSQ	797
Qy	60	CI-----NGVCWIVYHGAGTRTIAAPKGPVIQMYTNVDKDL-----VGWPAP	101
Db	798	ILKEHPTDNQELRNNRLYYLQNGK-----DMYTGODLDIHNLNSYDIDHIYP	846
Qy	102	QGSRLTPCTCGSSDLYLVTRHA-----DVIP---VRRRGD-----SRGSLISPRPIS	146
Db	847	QSF-----ITDNSTDNLVLTSSAGNREKGGDDVPPLEIVRKRKVFWEKLYOGNLMKRKFD	901
Qy	147	YL-KGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAV	180
Db	902	YLTRAERG--LTHADRARFIHROLVETROIITKNV	934

Search completed: August 30, 2003, 19:20:26
Job time : 18.0488 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 18:01:52 ; Search time 9.65768 seconds
(without alignments)
949.524 Million cell updates/sec

Title: US-09-965-594-12

Perfect score: 1021

Sequence: 1 MKKKGSVIVGRVINGAYA.....VAKAVDFIPVESLETTMRSP 195

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	894.5	87.6	3011	1 POLG_HCV1	P26664 h genome po
2	879.5	86.1	3011	1 POLG_HCVH	P27958 h genome po
3	847	83.0	3010	1 POLG_HCVTW	P29846 h genome po
4	842	82.5	3010	1 POLG_HCVBK	P26663 h genome po
5	838	82.1	3010	1 POLG_HCVJA	P26662 h genome po
6	837	82.0	3010	1 POLG_HCVJT	Q00269 h genome po
7	670.5	65.7	3033	1 POLG_HCVJ6	P26650 h genome po
8	668.5	65.5	3033	1 POLG_HCVJ8	P26661 h genome po
9	85.5	8.4	485	1 Y136_IIEPA	Q83172 treponema p
10	82.5	8.1	321	1 HHOA_ARATH	Q9sel7 arabidopsis
11	82	8.0	452	1 AAMP_HUMAN	Q13685 homo sapien
12	80.5	7.9	437	1 DEGL_ARATH	Q22609 arabidopsis
13	79	7.7	259	1 IRL1_HUMAN	P08833 homo sapien
14	78.5	7.7	209	1 PAAD_PSEAE	Q9hx08 pseudomonas
15	77.5	7.6	642	1 ENV_FLVGL	P08359 feline leuk
16	77.5	7.6	2663	1 CENE_HUMAN	Q02224 homo sapien
17	75.5	7.4	861	1 P058_CAEEL	P34552 caenorhabdi
18	75	7.3	639	1 ENV_FLVSA	P06752 feline leuk
19	74.5	7.3	415	1 2P1_RABIT	P48833 oryctolagus
20	74.5	7.3	776	1 HYPF_AZOVI	P40596 acetobacter
21	74	7.2	402	1 RAGE_RAT	Q63495 rattus norv
22	74	7.2	603	1 ENV_RSVP	P03396 rous sarcom
23	73.5	7.2	263	1 GRAK_MOUSE	Q35205 mus musculu
24	73.5	7.2	436	1 ENV_FLVCS	Q02077 feline leuk
25	73.5	7.2	661	1 INV8_DAUCA	P80065 daucus caro
26	73.5	7.2	1165	1 POL_GALV	P21414 gibbon ape
27	73	7.1	253	1 CAC3_BOVIN	P05805 bos taurus
28	73	7.1	645	1 ENV_FSVSM	P21445 feline sarc
29	73	7.1	676	1 ENV_MLVFP	P26803 friend muri
30	73	7.1	1705	1 PTPO_MOUSE	P70289 mus musculu
31	73	7.1	3414	1 POLG_LANVT	P29837 l genome po
32	72.5	7.1	257	1 GRAM1_HUMAN	P51124 homo sapien
33	72	7.1	659	1 VST2_REVME	Q03500 hepatitis e

RESULT 1
POLG_HCV1 STANDARD: PRT; 3011 AA.
AC P26664;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin) (EC 3.4.21.-); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate 1) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.
OX NCBI_TaxID=11104;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91172826; PubMed=1848704;
RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C., Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J., Bradley D.W., Kuo G., Houghton M.,
RA "Genetic organization and diversity of the hepatitis C virus.";
Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).
RL -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + (RNA)(N).
CC -1- SUMMARY: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M62321; AAA45676.1; -;
DR PIR; A39166; GNMVC3.
DR PDB; 1A1V; 16-FEB-99.
DR PDB; 1HEI; 25-NOV-98.
DR MEROPS; S29.001; -;
DR MEROPS; U39.001; -;
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.

P26804 friend muri
P39061 mus musculu
P11033 mus musculu
Q90627 gallus gall
P10205 herpes simp
P09855 herpes simp
P97608 rattus norv
P03221 epstein-bar
P11261 feline leuk
P54748 rattus norv
P12641 bovine herp
P18614 rattus norv

34 72 7.1 676 1 ENV_MLVFF
35 72 7.1 1527 1 CALH_MOUSE
36 71.5 7.0 248 1 GRAD_MOUSE
37 71.5 7.0 248 1 TRY1_CHICK
38 71.5 7.0 535 1 UL21_HSV11
39 71.5 7.0 535 1 UL21_HSV1E
40 71.5 7.0 1288 1 OPLA_RAT
41 71 7.0 336 1 UL16_EBV
42 71 7.0 662 1 ENV_FLVLB
43 71 7.0 844 1 CNMA_RAT
44 71 7.0 917 1 VGLB_HSVB2
45 71 7.0 1180 1 ITAL_RAT

ALIGNMENTS

DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002888; HCV_NS4c.
 DR InterPro: IPR002166; HCV_RdRP.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; helicase; 1.
 DR Pfam: PF00998; Viral_RdRP; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR PolyProtein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4 (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4b (POTENTIAL).
 FT CHAIN 2014 3011 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT CHAIN 347 369 POTENTIAL.
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH_BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 476 476 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 3011 AA: 327197 MW: 65F8C9447FCE5AF9 CRC64:
 Query Match 87.6%; Score 894.5; DB 1; Length 3011;
 Best Local Similarity 85.8%; Pred. No. 2.2e-76;
 Matches 175; Conservative 8; Mismatches 10; Indels 11; Gaps 2;
 3 KKGSWIVG---RIVLNG-----AYAOQTRGEGCOETSGTRKKNQVGEVQIVST 51

Db 1005 RRGREILLGPADGMVSKGWRLLAPITATYAQQRGLGCIITSLTGRKNQVGEVQIVST 1064
 Qy 52 AAQTFLATCINGVCWTVYHGAGTRTIASPKGVIOYMTNVKDLVGPAPQCSRLTPCT 111
 Db 1065 AAQTFLATCINGVCWTVYHGAGTRTIASPKGVIOYMTNVKDLVGPAPQCSRLTPCT 1124
 Qy 112 CGSSDLYLVTRHADVIPVRRGDSRGSLLSPRPISYLYKGSSGGPLLCFAGHAGVIFRAV 171
 Db 1125 CGSSDLYLVTRHADVIPVRRGDSRGSLLSPRPISYLYKGSSGGPLLCFAGHAGVIFRAV 1184
 Qy 172 CTRGKAVKAVDFIPVESLETTMRSP 195
 Db 1185 CTRGKAVKAVDFIPVENLETTMRSP 1208
 RESULT 2
 POLG_HCVH STANDARD: PRT: 3011 AA.
 AC P27958;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4a (P4); Nonstructural protein
 DE NS4b (P27); Nonstructural protein NS5a (P56); Nonstructural protein
 DE NS5b (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate H) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=11108;
 RP [1]
 RP SEQUENCE FROM N.A.
 RP MEDLINE-92052256; PubMed-1658800;
 RP Inchauspe G., Zebadee S., Lee D.H.H., Sugitani M., Nasoff M.,
 RA Prince A.M.;
 RT *Genomic structure of the human prototype strain H of hepatitis C
 RT virus: comparison with American and Japanese isolates.*;
 RT Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
 RL [2]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
 RP MEDLINE-9731322; PubMed-9187654;
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
 RT *Structure of the hepatitis C virus RNA helicase domain*;
 RT Nat. Struct. Biol. 4:463-467(1997).
 RL [3]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
 RP MEDLINE-98154321; PubMed-9493270;
 RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
 Murcko M.A., Lin C., Caron P.R.;
 RT *Hepatitis C virus NS3 RNA helicase domain with a bound
 RT oligonucleotide: the crystal structure provides insights into the mode
 RT of unwinding*;
 RL Structure 6:89-100(1998).
 CC -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
 CC -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
 CC -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
 CC ACTIVATION OF NS3.
 CC -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
 CC -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the p6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 CC (RNA)(N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.

CC -!- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
 CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
 CC -!- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
 CC -!- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M67463; AAA4534.1; .
 DR PIR: A36814; GNMVCH.
 DR PDB: 1HEI; 25-NOV-98.
 DR PDB: 1AIV; 16-FEB-99.
 DR PDB: 1A1R; 17-JUN-98.
 DR MEROPS: S29.001; .
 DR MEROPS: U39.001; .
 DR TRANSFAC: T04155; .
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV capsid.
 DR InterPro: IPR002521; HCV core.
 DR InterPro: IPR002519; HCV env.
 DR InterPro: IPR002531; HCV NS1.
 DR InterPro: IPR002518; HCV NS2.
 DR InterPro: IPR004109; HCV NS3.
 DR InterPro: IPR000745; HCV NS4a.
 DR InterPro: IPR001490; HCV NS4b.
 DR InterPro: IPR002868; HCV NS5a.
 DR InterPro: IPR002166; HCV RdRp.
 DR InterPro: IPR001450; Helicase_C.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSwir.
 DR Pfam: PF01543; HCV capsid; 1.
 DR Pfam: PF01542; HCV core; 1.
 DR Pfam: PF01539; HCV env; 1.
 DR Pfam: PF01560; HCV NS1; 1.
 DR Pfam: PF01538; HCV NS2; 1.
 DR Pfam: PF02907; HCV NS3; 1.
 DR Pfam: PF01006; HCV NS4a; 1.
 DR Pfam: PF01001; HCV NS4b; 1.
 DR Pfam: PF01506; HCV NS5a; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRp; 1.
 DR ProDom: PD186062; HCV NS1; 1.
 DR SMART: SM00487; DEXDc; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3d-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 1 191 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 192 383 ENVELOPE GLYCOPROTEIN E1.
 FT CHAIN 384 746 ENVELOPE GLYCOPROTEIN E2.
 FT CHAIN 747 809 PROTEIN P7.
 FT CHAIN 810 1026 NONSTRUCTURAL PROTEIN NS2.
 FT CHAIN 1027 1657 PROTEASE/HELICASE NS3.
 FT CHAIN 1658 1711 NONSTRUCTURAL PROTEIN NS4A.
 FT CHAIN 1712 1972 NONSTRUCTURAL PROTEIN NS4B.
 FT CHAIN 1973 2420 NONSTRUCTURAL PROTEIN NS5A.
 FT CHAIN 2421 3011 NONSTRUCTURAL PROTEIN NS5B.
 FT TRANSMEM 347 369 POTENTIAL.
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1224 1226
 FT TURN 1232 1233
 FT TURN 1236 1238
 FT HELIX 1239 1246
 FT TURN 1247 1248
 FT STRAND 1251 1255
 FT HELIX 1258 1271
 FT TURN 1272 1272
 FT STRAND 1277 1280
 FT TURN 1281 1282
 FT STRAND 1283 1285
 FT STRAND 1291 1295
 FT HELIX 1296 1301
 FT TURN 1302 1303
 FT STRAND 1312 1316
 FT TURN 1317 1319
 FT HELIX 1323 1335
 FT TURN 1336 1340
 FT STRAND 1343 1347
 FT TURN 1352 1353
 FT TURN 1360 1361
 FT STRAND 1362 1366
 FT STRAND 1368 1368
 FT STRAND 1373 1375
 FT TURN 1376 1377
 FT STRAND 1378 1380
 FT HELIX 1382 1385
 FT STRAND 1389 1393
 FT HELIX 1397 1409
 FT TURN 1410 1411
 FT STRAND 1414 1417
 FT TURN 1419 1420
 FT STRAND 1432 1436
 FT TURN 1438 1439
 FT STRAND 1450 1453
 FT STRAND 1456 1463
 FT STRAND 1471 1478
 FT STRAND 1480 1480
 FT HELIX 1481 1488
 FT TURN 1489 1490
 FT STRAND 1497 1501
 FT STRAND 1507 1507
 FT STRAND 1511 1511
 FT HELIX 1514 1527
 FT STRAND 1532 1544
 FT STRAND 1550 1550
 FT HELIX 1555 1564
 FT STRAND 1570 1578
 FT TURN 1579 1580
 FT HELIX 1584 1597
 FT TURN 1598 1598
 FT TURN 1606 1611
 FT TURN 1614 1618
 FT STRAND 1622 1623
 FT STRAND 1627 1627
 FT STRAND 1635 1636
 FT HELIX 1640 1652
 SQ SEQUENCE 3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;

Query Match 86.1%; Score 879.5; DB 1; Length 3011;
 Best Local Similarity 83.8%; Pred. No. 5.9e-75;

DR	InterPro	IPR002522	HCV_capsid.
DR	InterPro	IPR002521	HCV_core.
DR	InterPro	IPR002519	HCV_env.
DR	InterPro	IPR002531	HCV_NS1.
DR	InterPro	IPR002518	HCV_NS2.
DR	InterPro	IPR004109	HCV_NS3.
DR	InterPro	IPR000745	HCV_NS4a.
DR	InterPro	IPR001490	HCV_NS4b.
DR	InterPro	IPR002868	HCV_NS5a.
DR	InterPro	IPR002166	HCV_RdRp.
DR	InterPro	IPR007095	RNA_pol_DS_PS.
DR	InterPro	IPR007094	RNA_pol_PSVir.
DR	Pfam	PF01543	HCV_capsid; 1.
DR	Pfam	PF01542	HCV_core; 1.
DR	Pfam	PF01539	HCV_env; 1.
DR	Pfam	PF01560	HCV_NS1; 1.
DR	Pfam	PF01538	HCV_NS2; 1.

DK pfam: PF01006; HCV_NS4a, 1.
DR pfam: PF01001; HCV_NS4b, 1.

DR	Pfam:	PF01506;	HCV_NS5a; 1.
DR	Pfam:	PF00277;	Helicase_C; 1.
DR	Pfam:	PF00998;	Viral_HoRP; 1.
DR	ProDom:	PD186062;	HCV_NS1; 1.
DR	SMART:	SM00487;	DEDC; 1.
KW	Polypeptin:	Glycoprotein;	Transferase; RNA-directed RNA polymerase;
KW	Core protein;	Coat protein;	Envelope protein; Helicase; ATP-binding;
KW	Transmembrane;	Nonstructural protein;	Hydrolase; Serine protease;
KW	3D-structure.		
FT	INIT_MET	1	REMOVED FROM CAPSID PROTEIN C BY THE CELLULAR AMINOPEPTIDASE.
FT	CHAIN	1	CORE PROTEIN (POTENTIAL).
FT	CHAIN	116	MATRIX PROTEIN (POTENTIAL).
FT	CHAIN	192	MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT	CHAIN	384	NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT	CHAIN	729	NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT	CHAIN	1006	PROTEASE/HELICASE NS3 (POTENTIAL).
FT	CHAIN	1615	NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT	CHAIN	1616	NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT	CHAIN	1963	NONSTRUCTURAL PROTEIN NS4C (POTENTIAL).
FT	CHAIN	2013	RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT	CHAIN	2014	POTENTIAL.
FT	TRANSMEM	347	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	1083	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	1107	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	1165	ATP (POTENTIAL).
FT	NP_BIND	1230	DECH BOX.
FT	SITE	1316	N-LINKED (GLCNAC..)
FT	CARBOHYD	196	N-LINKED (GLCNAC..)
FT	CARBOHYD	209	N-LINKED (GLCNAC..)
FT	CARBOHYD	233	N-LINKED (GLCNAC..)
FT	CARBOHYD	234	N-LINKED (GLCNAC..)
FT	CARBOHYD	250	N-LINKED (GLCNAC..)
FT	CARBOHYD	305	N-LINKED (GLCNAC..)
FT	CARBOHYD	417	N-LINKED (GLCNAC..)
FT	CARBOHYD	423	N-LINKED (GLCNAC..)
FT	CARBOHYD	430	N-LINKED (GLCNAC..)
FT	CARBOHYD	448	N-LINKED (GLCNAC..)
FT	CARBOHYD	532	N-LINKED (GLCNAC..)
FT	CARBOHYD	540	N-LINKED (GLCNAC..)
FT	CARBOHYD	556	N-LINKED (GLCNAC..)
FT	CARBOHYD	576	N-LINKED (GLCNAC..)
FT	CARBOHYD	623	N-LINKED (GLCNAC..)
FT	CARBOHYD	645	N-LINKED (GLCNAC..)
FT	CARBOHYD	2041	N-LINKED (GLCNAC..)
FT	CARBOHYD	2077	N-LINKED (GLCNAC..)
FT	CARBOHYD	2240	N-LINKED (GLCNAC..)
FT	CARBOHYD	2529	N-LINKED (GLCNAC..)
FT	CARBOHYD	2788	N-LINKED (GLCNAC..)
FT	CARBOHYD	2788	N-LINKED (GLCNAC..)
SEQ	SEQUENCE	3010 AA;	AAD267D55CDFE215 CRC64;

Query Match 83.0%; Score 847; DB 1; Length 3010;
 Best Local Similarity 88.2%; Pred. No. 7.le-72;
 Matches 157; Conservative 11; Mismatches 10; Indels 0; Gaps

Query Match 83.0%; Score 847; DB 1; Length 3010;
Best Local Similarity 88.2%; Pred. No. 7.le-72;
Matches 157; Conservative 11; Mismatches 10; Indels

QY 18 AYAOOTRGECCQETSOTGRKNOVEGEVOIVSTAAOTFLATCINGCVTVYHGAGTRII 77
 DB 1031 AYAOOTRGI.FGCIITSLTGRDKNOVEGEVQVSTATOSFLATCINGCVTVYHGAGSKTL 1090
 QY 78 ASYKGPVIMYTNVDKDLVWPAPOGSKSLPTCTCGSSDLYLVTRHADVLPVRRRGDSRG 137
 DB 1091 AGPKGPITOMYTNVDQDLVWHPAQARSLLPTCTCGSSDLYLVTRHADVLPVRRRGDSRG 1150
 QY 138 SLLSPRPISYLKSGSGGGLPCAGHAGVIFPAAACTRGVAKAVDFIPVESLETTMRSP 195
 DB 1151 SLLSPRPISYLKSGSGGGLPCGHRVGVIFPAAVCTRGVAKAVDFIPVESMETMRSP 1208

RESULT 4

POLG_HCVBK
 ID POLG_HCVBK STANDARD; PRT: 3010 AA.
 AC F26663;
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (Isolate BK) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11105;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91140698; PubMed=1847440;
 RA Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
 RA Onishi E., Andoh T., Yoshida I., Okayama H.;
 RT "Structure and organization of the hepatitis C virus genome isolated
 RT from human carriers.";
 RL J. Virol. 65:1105-1113(1991).
 RN [2]
 RP SEQUENCE OF 1487-1500.
 RX MEDLINE=96235224; PubMed=8647104;
 RA Borowski P., Heiland M., Oehlmann K., Becker B., Kornetevy L.;
 RT "Non-structural protein 3 of hepatitis C virus inhibits
 RT phosphorylation mediated by cAMP-dependent protein kinase.";
 RL Eur. J. Biochem. 237:611-618(1996).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
 RX MEDLINE=97015088; PubMed=8861916;
 RA Love R.A., Parge H.E., Wickersham J.A., Hostomsky Z., Habuka N.,
 RA Moosaw E.W., Adachi T., Hostomska Z.;
 RT "The crystal structure of hepatitis C virus NS3 proteinase reveals a
 RT trypsin-like fold and a structural zinc binding site.";
 RL Cell 87:331-342(1996).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
 RX MEDLINE=98227846; PubMed=9568891;
 RA Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
 RA Steinkuehler C., Tomei L., de Francesco R., Kuo I.C., Chen Z.;
 RT "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C
 RT virus: a 2.2-A resolution structure in a hexagonal crystal form.";
 RL Protein Sci. 7:837-847(1998).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate ~ N diphosphate +
 CC [RNA](N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

CC CC PROTEIN C AND MRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: M58335; GAA72945.1;
 CC PIR: A38465; GAAVTC.
 CC DR PDB: 1AJQ; 25-MAR-98.
 CC DR PDB: 1JXP; 14-JAN-98.
 CC DR PDB: 1NS3; 08-APR-98.
 CC DR PDB: 1C2P; 15-NOV-00.
 CC DR PDB: 1CSJ; 08-NOV-99.
 CC DR PDB: 1GX5; 09-APR-02.
 CC DR PDB: 1GX6; 10-APR-02.
 CC DR PDB: 1QVJ; 26-JUN-00.
 CC DR PDB: 8OHM; 20-APR-99.
 CC DR MEROPS: S29.001;
 CC MEROPS: U39.001;
 CC DR InterPro: IPR001410; DEAD.
 CC DR InterPro: IPR002522; HCV_capsid.
 CC DR InterPro: IPR002521; HCV_core.
 CC DR InterPro: IPR002519; HCV_env.
 CC DR InterPro: IPR002531; HCV_NS1.
 CC DR InterPro: IPR002518; HCV_NS2.
 CC DR InterPro: IPR004109; HCV_NS3.
 CC DR InterPro: IPR000745; HCV_NS4a.
 CC DR InterPro: IPR001490; HCV_NS4b.
 CC DR InterPro: IPR002868; HCV_NS5a.
 CC DR InterPro: IPR002166; HCV_RDRP.
 CC DR InterPro: IPR007095; RNA_pol_DS_PS.
 CC DR InterPro: IPR007094; RNA_pol_PSVir.
 CC DR Pfam: PF01543; HCV_core; 1.
 CC DR Pfam: PF01539; HCV_env; 1.
 CC DR Pfam: PF01560; HCV_NS1; 1.
 CC DR Pfam: PF01538; HCV_NS2; 1.
 CC DR Pfam: PF02907; HCV_NS3; 1.
 CC DR Pfam: PF02006; HCV_NS4a; 1.
 CC DR Pfam: PF01001; HCV_NS4b; 1.
 CC DR Pfam: PF01506; HCV_NS5a; 1.
 CC DR Pfam: PF00998; Viral_RDRP; 1.
 CC DR ProDom: PD186062; HCV_NS1; 1.
 CC DR SMART: SM00487; DEXDC; 1.
 CC DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 CC Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 CC Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 CC 3D-structure.
 CC INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 CC CELLULAR AMINOPEPTIDASE.
 CC CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).
 CC CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
 CC CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 CC CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 CC CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 CC CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
 CC CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 CC CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 CC CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 CC TRANSMEM 347 369 POTENTIAL.
 CC ACT_SITE 1083 1083 CHARGE RELAY SYSTEM.
 CC ACT_SITE 1107 1107 CHARGE RELAY SYSTEM.
 CC ACT_SITE 1165 1165 CHARGE RELAY SYSTEM.
 CC NP_BIND 1230 1237 ATP (POTENTIAL).
 CC SITE 1316 1319 DECH BOX.
 CC FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).


```

DR Pfam: PF01538; HCV_NS2: 1.
DR Pfam: PF02907; HCV_NS3: 1.
DR Pfam: PF01006; HCV_NS4a: 1.
DR Pfam: PF01003; HCV_NS4b: 1.
DR Pfam: PF01506; HCV_NS5a: 1.
DR Pfam: PF00271; Helicase_C: 1.
DR Pfam: PF00998; Viral_RDRP: 1.
DR Pfam: PF0186062; HCV_NS1: 1.
DR SMART: SM00487; DEXoc: 1.
KW Polyprotein: Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
FT INIT_MET 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 729
FT CHAIN 730 1006
FT CHAIN 1007 1615
FT CHAIN 1616 1862
FT CHAIN 1863 2013
FT CHAIN 2014 3010
FT CHAIN 3010 347
FT TRANSMEM 347 369
FT ACT_SITE 1083 1083
FT ACT_SITE 1107 1107
FT ACT_SITE 1165 1165
FT NP_BIND 1230 1237
FT SITE 1316 1319
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 250 250
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 532 532
FT CARBOHYD 556 556
FT CARBOHYD 576 576
FT CARBOHYD 623 623
FT CARBOHYD 645 645
FT CARBOHYD 2041 2041
FT CARBOHYD 2077 2077
FT CARBOHYD 2240 2240
FT CARBOHYD 2788 2788
FT SEQUENCE 3010 AA: 327017 MW: AA993794F46DB185 CRC64;
Query Match 82.1%; Score 838; DB 1; Length 3010;
Best Local Similarity 85.4%; Pred No. 5, le-71;
Matches 152; Conservative 16; Mismatches 10; Indels 0; Gaps 0;
OY 18 AYAOOTRGEQCQETSGTRDKNQVEGVQIVSTAAQTFFLATCINGVGTWVYHCAGTETI 77
Db 1031 AYSOOTRGLGCIITSLGRDKNQVDGEVQLSTATQSLFATCVGVGTWVYHCAGSKTL 1090
OY 78 ASPKGPVIOYTNVDKDLVGHVPAQGSRLTPTCTCGSSDLVLTTHADVIVPRRGDSRG 137
Db 1091 AGKGPITOMYTNVDUULVGPAPPAGKSMPTCTCGSSDLVLTTHADVIVPRRGDSRG 1150
OY 138 SLLSPRISYLGSGGGPLLCFAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMRSP 195
Db 1151 SLLSPRISYLGSGGGPLLCFSGHVGFIRAVCTRGVAKAVDFIPVESLETTMRSP 1208
RESULT 6
ID POLG_HCVJT STANDARD: PRT: 3010 AA.
AC Q00269;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)

```

```

DE DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-JT) (HCV)
OS Viruses; sRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID:31642;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:92295714; PubMed:1318627;
RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,
RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
RT "Molecular cloning of hepatitis C virus genome from a single Japanese
RT carrier: sequence variation within the same individual and among
RT infected individuals.";
RL Virus Res. 23:39-53(1992).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
CC [RNA](N).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC PMBL: D11168; BAA01943.1; -
DR PIR: A45573; A45573.
DR PDB: 1AIQ; 25-MAR-98.
DR PDB: 1JXP; 14-JAN-98.
DR MEROPS: S29.001; -
DR MEROPS: U39.001; -
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4b.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RDRP.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; Helicase_C; 1.
DR Pfam: PF00998; Viral_RDRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.

```

SMART; SM00487; DEXdc; 1.
 KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CELLULAR AMINOPEPTIDASE.
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT CHAIN 3011 369
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 532 532
 FT CARBOHYD 540 540
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2529 2529
 FT CARBOHYD 2788 2788
 FT SEQUENCE 3010 AA; 326573 MW; 94ALC77435D642BB CRC64;
 Query Match 82.0%; Score 837; DB 1; Length 3010;
 Best Local Similarity 86.5%; Pred. No. 6,3e-71;
 Matches 154; Conservative 13; Mismatches 11; Indels 0; Gaps 0;
 QY 18 AYAOOTRCEEGCOETSTGRDKNOVEGEVOIVSTAAOTFLATCINGVCVTYHGAGTRTI 77
 Db 1031 AYAOOTRGLGCVITSLTGRDKNOVEGEVOVSTATQSFATCNGVCVTYHFGAGSKTL 1090
 QY 78 ASPKGPVITOMYTNVDKLVGWPAFGSGSLTPCTCGSSDLYLVTRHADVTPVRRRDSRG 137
 Db 1091 AGPKGPITOMYTNVDQDLVGHHPAGSRLTPCTCGSSDLYLVTRHADVTPVRRRDSRG 1150
 QY 138 SLSPRPISYLGSGGPGLLCPAGHAGVLEPAACVTRGVAKAVDFIPVESLETTMSP 195
 Db 1151 SLSPRPISYLGSGGPGLLCPGSHAVCIFFRAAVCTRGVAKAVDFIPVESNETTMSRP 1208
 RESULT 7
 ID POLG_HCVJ6 STANDARD; PRT: 3033 AA.
 AC P26660;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polypeptide [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirus)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein

DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-J6) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11113;
 RN [J]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92044440; PubMed=1658196;
 RA Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H.,
 RA Machida A., Miyakawa Y., Mayumi M.;
 RT "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
 from a human carrier: comparison with reported isolates for conserved
 and divergent regions.";
 RT J. Gen. Virol. 72:2697-2704 (1991).
 RL J. Gen. Virol. 72:2697-2704 (1991).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 precursor polypeptide, commonly with Asp or Glu in the P6
 position, Cys or Thr in P1 and Ser or Ala in P1',
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 [RNA](N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL; D00944; BAA00792.1; -;
 DR PIR; JQ1303; JQ1303.
 DR HSSP; P27958; IHEI.
 DR MEROPS; S29.001; -;
 DR MEROPS; U39.001; -;
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR004109; HCV_NS3.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RdRP.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_Psvir.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00271; helicase_C; 1.
 DR Pfam; PF00998; viral_RdRP; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXdc; 1.
 KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
 FT INIT_MET 1 1
 REMOVED FROM CAPSID PROTEIN C BY THE

FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CARBOHYD 627 627
 FT CARBOHYD 649 649
 FT CARBOHYD 1091 1091
 FT CARBOHYD 2038 2038
 FT CARBOHYD 2811 2811
 SQ SEQUENCE 3033 AA; 329165 MW; F957F5C1A273BE9E CRC64;

Query Match 65.78; Score 670.5; DB 1; Length 3033;
 Best Local Similarity 62.19; Pred. No. 3,8e-55;
 Matches 126; Conservative 27; Mismatches 33; Indels 17; Gaps 1;

QY 10 VGRVLNG-----AYAQOTREEGCOETQSGTRKDNKQVEGVOIVSTA 52
 DB 1010 LGREVLGPADGYTSKWSLLAPITAYAQOTRGLLGTIVVSMGTGRDKTQAGSGIQLSTV 1069
 QY 53 AQTFELARCINGCVTVYHGAGRTIASPGVQIMYINVDKLVGVPAPGSRSLPCTC 112
 DB 1070 TQSEPLGTISGVLTVYHGAGNKTLAGSRGVPQIMYSSAEGDLVGMPSPGTKSLPCTC 1129
 QY 113 GSSDLVYVTRHADVIPVRHNCDSRGSLSPRISYHKGSSGGPILCPAGHAVGIFRAAVC 172
 DB 1130 GAVDLYLVTRNADVIPARRRDKRGALLSPRLSLKSGSGGPVLCPRHAVGVFRAAVC 1189
 QY 173 TRGVAKAVDFIPVESLETTMRSP 195
 DB 1190 SRGVAKSIDFIPVILDIVTRSP 1212

RESULT 8
 ID POLG_HCVJ8 STANDARD; PRT: 3033 AA.
 AC P26661;
 DT 01-AUG-1992 (Rel. 23, Last Created)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-J8) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.
 OX NCBI_TaxID=11115;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230332; PubMed=1314459;
 RA Okamoto H., Kural K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,
 RA Fukuda S., Tsuda F., Mishiro S.;
 RT *Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes.;
 RL Virology 188:331-341(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 CC [RNA](N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: D10988; BAA01761.1; -.
 CC PIR: A40250; GNMVJ8.
 CC HSSP: P27958; LHEI.
 CC MEROPS: S29.001; -.
 CC MEROPS: U39.001; -.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR007045; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RDRP.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00998; Viral_RDRP; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 192 733
 FT CHAIN 734 1010

FT CHAIN 1011 1619 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1620 1865 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 1867 2017 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 2018 3033 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT TRANSMEM 347 369 POTENTIAL.
 FT ACT_SITE 1087 1087 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1111 1111 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1169 1169 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1234 1241 ATP (POTENTIAL).
 FT SITE 1320 1323 DECH_BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 299 299 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 477 477 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 534 534 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 542 542 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 558 558 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 578 578 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 627 627 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 649 649 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1091 1091 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2038 2038 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2359 2359 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2811 2811 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 3033 AA; 16173E7E3381FDIA CHC64;
 Query Match 65.5%; Score 668.5; DB 1; Length 3033;
 Best Local Similarity 62.6%; Pred. No. 5.9e-55;
 Matches 127; Conservative 25; Mismatches 34; Indels 17; Gaps 1;
 QY 10 VGRVLVNG-----AYAQOTCEGCGQETSTQGRDKNOVEGVEQIVSTA 52
 DB 1010 LGREVLGPADGYTSKGWKLAIATAYTQTRGLLGAIVVSLTRDKNOAGQOVQLSSV 1069
 QY 53 AQTFLATCINGCVTVYHGAGRTIASPKGPVIQMTYNDKDLVGPAPQGSRSSTPCTC 112
 DB 1070 TOTELGTSISGLVTVYHGAGNCTLAGPKGPVIMTYSAEGDLVGVPPSGTKSLDPCTC 1129
 QY 113 GSSDLXLYTRHADVPVRRRGDSRGLSPRPISYLSKSGGGLCPAGHAGVIFRAVC 172
 DB 1130 GAVDLYLVRNADVPVRRDRRGLSPRLTLKSGSGGVPVLCRSHAGVGLFRAVC 1189
 QY 173 TRGVAKAVDFIPVESLETTMRSP 195
 DB 1190 ARGVAKSIDFIPVESLVDVATRP 1212

RESULT 9

Y136_TREPA
 ID Y136_TREPA STANDARD; PRT; 485 AA.
 AC O83172;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical lipoprotein TP0136 precursor.
 GN TP0136.
 OS Treponema pallidum.
 OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
 ON NCBI_TaxID=160;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Nichols;
 RX MEDLINE=98332770; PubMed=36656576;
 RA Fraser C.M., Norris S., Weinstock G.M., White O., Sutton G.G.,
 RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
 RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
 RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,

RA McDonald L., Attiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
 RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
 RA Venter J.C.;
 RT "Complete genome sequence of Treponema pallidum, the syphilis
 RT Spirochete";
 RL Science 281:375-388(1998).
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
 CC (POTENTIAL).
 CC -!- SIMILARITY: BELONGS TO THE TP0136 FAMILY OF LIPOPROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AE001199; AAC65137.1; ALT_INIT.
 CC TIGR; TP0136; .
 DR Hypothetical protein; Lipoprotein; Membrane; Signal;
 KW Complete proteome.
 FT SIGNAL 1 23 POTENTIAL.
 FT CHAIN 24 485 HYPOTHETICAL LIPOPROTEIN TP0136.
 FT LIPID 24 24 N-ACYL DIGLYCERIDE (POTENTIAL).
 FT DOMAIN 164 178 GLY/SER-RICH.
 FT DOMAIN 196 210 GLY/SER-RICH.
 FT DOMAIN 253 267 GLY/SER-RICH.
 FT DOMAIN 318 327 POLY-SER.
 FT DOMAIN 444 447 POLY-SER.
 SQ SEQUENCE 485 AA; 48984 MW; C7A4CEEDC7DC5CED CRC64;

Query Match 8.4%; Score 85.5; DB 1; Length 485;
 Best Local Similarity 23.6%; Pred. No. 1.3;
 Matches 52; Conservative 19; Mismatches 78; Indels 71; Gaps 11;
 QY 3 KKGSVIV--GRVLNGAYAAQOTRGECCQETSO----TGRDKNOVEGVEQIVSTAQTFF 56
 DB 53 KAGSKLYATNGRL-----WEKELNGTSGWOKVSSSSVPTDSK-----KVMSTATDNTFF 102
 QY 57 LATCI--NGVCWTYVYHGAG---TRTIASPKGPVIQMTYNDKDLVG-----WPAQGSR 105
 DB 103 VLACVPGTGVYKHCNVAGSGSSSTGTATSPSTECQSOHAT---LVGTSKPEWLVPGGTF 158
 QY 106 SLTPCTC-----GSSDLVLYTRHADVP-----VRRRGDSRGLSPRPISYLSK--- 149
 DB 159 NNGMCGGGGGGGSSSSSCIIHLVLPVGTGNNGCGGGGGGGSSSSSCIIHKIVEN 218
 QY 150 -----GSSGGPLLCAGHAGV 165
 DB 219 TDEQFLDMGEGYVYVTKHLYTKNGSSSAGPAQCPCGGGGG 258

RESULT 10

HHOA_ARATH
 ID HHOA_ARATH STANDARD; PRT; 321 AA.
 AC Q98EL7; O49507;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Protease HhoA, chloroplast precursor (EC 3.4.21.-).
 GN HHOA OR AT4G18370 OR F28J12.30.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 ON NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lensch M.H.A., Sokolenko A., Herrmann R.G.;
 RT "Identification and characterization of the chloroplast HhoA protease,
 RT a homolog to the bacterial periplasmic protease HhoA";
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.

[2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv Columbia;
 RX MEDLINE=20083488; PubMed=10617199;
 RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
 RA Pohl T., Dueterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
 RA Harris B., Ansoorge W., Brandt P., Grivell L., Rieger M.,
 RA Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
 RA Kreis M., Delsen M., Puigdomenech P., Watson M., Schmidheini T.,
 RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
 RA Vos P., Hohenseil J., Zimmermann W., Wedler H., Ridley P.,
 RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
 RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
 RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
 RA Weitzenecker T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
 RA Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
 RA Moeljan P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
 RA Bernetser S., Hempel S., Feldpausch M., Lamberth S., van den Daelc H.,
 RA De Keyser A., Buysshaert C., Gielen J., Villarroel R., De Clercq R.,
 RA Van Montagu M., Kogers J., Cronin A., Quail M., Gray-Allen S.,
 RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,
 RA Pettett A., Rajandream M.A., Lyne M., Benes V., Reclmann S.,
 RA Borkova D., Bloeker H., Scharfe M., Grimm M., Loehner I.-H.,
 RA Dose S., de Haan M., Maarse A., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fartmann B., Grandrath K., Dauner D., Herzl A.,
 RA Neumann S., Argirou F., Vitale D., Liguori R., Piravandi F.,
 RA Massenot O., Quigley F., Clabaud G., Muendlein A., Felber R.,
 RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
 RA Chefdor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
 RA Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,
 RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse I.,
 RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bieleke C.,
 RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
 RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
 RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
 RA Senkon M., Murray J., Sheet P., Cordes M., Abu-Threiden J.,
 RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
 RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
 RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
 RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
 RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
 RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
 RA Antonoli B., Zidanic M., Strong C., Sun H., Lamar B., Jordan C.,
 RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
 RA Granat S., Shoddy N., Hasegawa A., Rameed A., Lodhi M., Johnson A.,
 RA Chen E., Marra M., Martienssen R., McCombie W.R.,
 RA "Sequence and analysis of chromosome 4 of the plant Arabidopsis
 RI thaliana.";
 RL Nature 402:769-777(1999).
 RN [3]
 RP SEQUENCE OF 72-82: 96-110: 150-159: 178-211 AND 306-320.
 RA Schubert M., Peterson U., Funk C., Haas B., Schroeder W.P.,
 RA Kieselbach T.,
 RT The chloroplast lumen from Arabidopsis thaliana.";
 RL Submitted (JUL-2001) to the SWISS-PROT data bank.
 CC -!- SUBCELLULAR LOCATION: Chloroplast; within the thylakoid lumen.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
 CC -!- CAUTION: Ref.2 sequences differ from that shown due to erroneous
 CC gene model prediction. AT4G18370 and AT4G18375 were originally
 CC fused into a single gene.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AF114386; AAF24060.1;
 CC EMBL: AL021710; CAA16717.1; ALT_SEQ.
 CC EMBL: AL161548; CAB78839.1; ALT_SEQ.
 CC
 CC DR EMBL: AF114386; AAF24060.1;
 CC DR EMBL: AL021710; CAA16717.1; ALT_SEQ.
 CC DR EMBL: AL161548; CAB78839.1; ALT_SEQ.

DR MEROPS: S01.279;
 DR InterPro: IPR001940; Protease2C.
 DR InterPro: IPR001254; Ser_protease_Try.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00834; PROTEASES2C.
 KW Hydrolase; Serine protease; Chloroplast; Thylakoid; Transit peptide.
 FT TRANSIT 1 26
 FT TRANSIT 27 71
 FT TRANSIT 72 321
 FT CHAIN 77 87
 FT DOMAIN 145 145
 FT ACT_SITE 186 186
 FT ACT_SITE 264 264
 FT ACT_SITE 40 40
 FT CONFLICT 321 AA: 34691 MW; 68DB81E0BD27A7A7 CRC64;
 FT SEQUENCE 321 AA: 34691 MW; 68DB81E0BD27A7A7 CRC64;
 Query Match 8.1%; Score 82.5; DB 1: Length 321;
 Best Local Similarity 26.2%; Pred. No. 1.6;
 Matches 39; Conservative 22; Mismatches 49; Indels 39; Gaps 8;
 QY 73 GRTTASPQGVIOYVTKVNDKLVGWPAPQSGSRSLTPTCGSSDLYLVTRHADVIPRRR 132
 DB 169 GTR--FSKGGKIVGL--DPDNDLAVLKIEGRELNPVVLGTSNDRVGQSCFAI----- 219
 QY 133 GDSRG-----SLLSPRPISYK-----GSSGGPLLCAGHAGVIF 167
 DB 220 GNPYGYENTLTIGVYSGLGREIPSPNGKISIAIOTDADINSNGSGPLDSYGHGTIGV- 278
 QY 168 RAACVTR---GVAKAVDF-IPVESLETTM 192
 DB 279 NTATPTKSGMSSGVNFAIDTIVRTV 307
 RESULT 11
 AAMP_HUMAN STANDARD; PRT: 452 AA.
 ID AAMP_HUMAN STANDARD; PRT: 452 AA.
 AC Q13685;
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DI 15-JUL-1998 (Rel. 36, Last sequence update)
 DE 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Angio-associated migratory cell protein.
 DE AAMP.
 GN Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Brain;
 RX MEDLINE=95262124; PubMed=7743515;
 RA Beckner M.E., Krutzsch H.C., Stracke M.L., Williams S.T.,
 RA Gallardo J.A., Liotta L.A.;
 RT Identification of a new heparin-binding superfamily protein expressed
 RI in blood vessels with a heparin-binding consensus sequence.";
 RL Cancer Res. 55:2140-2149(1995).
 CC -!- FUNCTION: MAY HAVE A FUNCTION IN MIGRATING CELLS.
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN BLOOD VESSELS. STRONGLY EXPRESSED
 CC IN ENDOTHELIAL CELLS, CYTOTROPHOBLASTS, AND POORLY DIFFERENTIATED
 CC COLON ADENOCARCINOMA CELLS FOUND IN LYMPHATICS.
 CC -!- SIMILARITY: Contains 8 WD repeats.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M95627; AAA68889.1;
 CC PIR: I39383; I39383.
 CC Genew: HGNC:18; AAMP.
 CC MIM: 603488; -.

DR GO:0008201; F:heparin binding activity; TAS.
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF00400; WD40; 8.
 DR SMART: SM00320; WD40; 8.
 DR PROSITE: PS00678; WD_REPEATS_1; 1.
 DR PROSITE: PS00082; WD_REPEATS_2; 6.
 DR PROSITE: PS0294; WD_REPEATS_REGION; 1.
 KW Repeat: WD repeat.
 FT DOMAIN 14 18 HEPARIN-BINDING (POTENTIAL).
 FT DOMAIN 71 77 POLY-GLU.
 FT REPEAT 107 138 WD 1.
 FT REPEAT 150 180 WD 2.
 FT REPEAT 190 220 WD 3.
 FT REPEAT 231 261 WD 4.
 FT REPEAT 276 306 WD 5.
 FT REPEAT 333 363 WD 6.
 FT REPEAT 374 404 WD 7.
 FT REPEAT 416 446 WD 8.
 SQ SEQUENCE 452 AA; 49015 MW; DA1413D25FE236C0 CRC64;

Query Match 8.0%; Score 82; DB 1; Length 452;
 Best Local Similarity 25.3%; Pred. No. 2.6;
 Matches 42; Conservative 13; Mismatches 47; Indels 64; Gaps 9;

QY 66 WTYYHGAGRTTIASPKGPVIMYTNVDKDLGVGWPAPQGSRL-----TPCTCGSSDLIV 120
 DB 197 WMEWH-----PRAPVLLAGT-ADGNTMMKVPNGDKTFQGNPCPATCGR----- 240
 QY 121 TRIADVIVRRR-----GDSRGS-----LLSPRISYKLGSSG--GPLICPA----- 160
 DB 241 -----VLPDGKRAVGVYEDGTIRIMDLKQSPHYLVKGTGEGHQPPLTCVAANQDGLILT 295
 QY 161 -----GHAVGIFR-----AAVCTRGVAKVDFIPVESL 188
 DB 296 GSYDCQAKLSVATTKGVGVGFETVATSPQSLGEGESESNSVESL 341

RESULT 12
 DEGI_ARATH STANDARD; PRT: 437 AA.
 AC O22609; Q9LX85;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Protease Do-like 1, chloroplast precursor (EC 3.4.21.-).
 GN DEPI OR DEPI OR A3G27923 OR K16N12.18.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RA MEDLINE=98175982; PubMed=9507020;
 RX Itzhaki H., Naveh I., Lindahl M., Cook M., Adam Z.;
 RT "Identification and characterization of DegP, a serine protease
 RT associated with the luminal side of the thylakoid membrane.";
 RL J. Biol. Chem. 273:7094-7098 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=20363099; PubMed=10907853;
 RA Kaneko T., Katoh T., Sato S., Nakamura A., Asamizu E., Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
 RT Sequence features of the 4,251,695 bp regions covered by 90 Pl. TAC
 RT and BAC clones.";
 RL DNA Res. 7:217-221 (2000).
 RN [3]
 RP SEQUENCE OF 104-118.
 RC STRAIN=cv. Columbia;
 RA Kieselbach T., Bystedt M., Schroeder W.P.;
 RL Submitted (JUL-2000) to the SWISS-PROT data bank.
 CC -!- FUNCTION: SERINE PROTEASE THAT IS REQUIRED AT HIGH TEMPERATURE.

MAY BE INVOLVED IN THE DEGRADATION OF DAMAGED PROTEINS. IN VIVO,
 CAN DEGRADE BETA-CASEIN.
 -!- ENZYME REGULATION: INHIBITED BY PHENYLMETHYLSULFONYL FLUORIDE AND
 O-PHENANTHROLINE.
 -!- SUBCELLULAR LOCATION: BOUND TO LUMINAL SIDE OF THE THYLAKOID
 MEMBRANE.
 -!- INDUCTION: By heat shock.
 -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
 -!- SIMILARITY: Contains 1 PDZ/DHR domain.

 This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).

 EMBL; AF028842; AAC39436.1; -;
 EMBL; AP000371; BAB02539.1; -;
 EMBL; AP001302; BAB02539.1; JOINED.
 MEROPS: S01.279; -;
 InterPro: IPR001478; PDZ.
 InterPro: IPR001940; Protease2C.
 InterPro: IPR001254; Ser_protease_Try.
 Pfam: PF00595; PDZ; 1.
 Pfam: PF00089; trypsin; 1.
 PRINTS: PR00834; PROTEASES2C.
 SMART: SM00228; PDZ; 1.
 PROSITE; PS0106; PDZ; 1.
 Hydrolyase; Serine protease; Transit peptide; Chloroplast; Thylakoid.
 FT TRANSIT 1 ?
 FT CHAIN 104 437 THYLAKOID.
 FT DOMAIN 152 321 PROTEASE DO-LIKE 1.
 FT DOMAIN 324 421 SERINE PROTEASE.
 FT ACT_SITE 171 171 PDZ.
 FT ACT_SITE 201 201 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 280 280 CHARGE RELAY SYSTEM (POTENTIAL).
 FT CONFLICT 12 23 HSPSSQLGNST -> SSTFLHSPPSHL (IN REF.
 2).
 FT CONFLICT 36 36 V -> I (IN REF. 2).
 FT CONFLICT 54 54 P -> S (IN REF. 2).
 FT CONFLICT 60 60 G -> R (IN REF. 2).
 FT CONFLICT 64 64 G -> D (IN REF. 2).
 FT CONFLICT 68 69 LL -> HF (IN REF. 2).
 FT CONFLICT 355 355 L -> V (IN REF. 2).
 FT CONFLICT 381 381 I -> V (IN REF. 2).
 FT CONFLICT 416 416 Q -> E (IN REF. 2).
 SQ SEQUENCE 437 AA; 46213 MW; 1497B1AB3F5FF2A4 CRC64;

Query Match 7.9%; Score 80.5; DB 1; Length 437;
 Best Local Similarity 25.6%; Pred. No. 3.5;
 Matches 44; Conservative 18; Mismatches 55; Indels 55; Gaps 7;

QY 68 VYHGAGRTTIASPKGPVIMY-----TNVDKDLGVG-----PA 100
 DB 150 VPQSGSGFFWMDKGGHIVTNYHVRGASDLRVTLADQTTFDKAVGFDQDKVAVLRDA 209
 QY 101 PQGSRSLTPCTCGSSDLIVL-----TRHADVIPVRRRSGSLLSPRT 145
 DB 210 PK--NKLRIPIVGVVSADLLVQKVFALGNPFGLDHTLTGTVISGLRREIS--SAATGRPI 265
 QY 146 SYL-----KGSSGGLPLCPAGHVGIFRAVCTRGVAKAVDF-IPVESL 188
 DB 266 QDVITQTDAAINPGSGGLDSSGTGLIGINTAIYSPSGASSGVGFSPVDTV 317

RESULT 13
 IBP1_HUMAN
 ID IBP1_HUMAN STANDARD; PRT: 259 AA.
 AC P08633;
 DT 01-NOV-1988 (Rel. 09, Created)

DT 01-NOV-1988 (rel. 09, last sequence update)
DT 28-FEB-2003 (rel. 41, last annotation update)
DE Insulin-like growth factor binding protein 1 precursor (IGFBP-1)
DE (IBP-1) (IGF-binding protein 1) (Placental protein 12) (PP12).
GN IGFBP1 OR IBP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=89052654; PubMed=2461294;
RA Brinkman A., Groffen C., Kortleve D.J., Geurts A., Drop S.L.S.;
RT "Isolation and characterization of a cDNA encoding the low molecular
RT weight insulin-like growth factor binding protein (IBP-1).";
RL EMBO J. 7:2417-2423(1988).
RN [2]
RN SEQUENCE FROM N.A.
RC TISSUE=Decidua;
RX MEDLINE=88240345; PubMed=2454104;
RA Brewer M.T., Stetler G.L., Squires C.H., Thompson R.C.,
RA Busby W.H. Jr., Clemmons D.R.;
RT "Cloning, characterization, and expression of a human insulin-like
RT growth factor binding protein.";
RL Biochem. Biophys. Res. Commun. 152:1289-1297(1988).
RN [3]
RN SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=88335621; PubMed=3419031;
RA Grundmann U., Nerlich C., Bohn H., Rein T.;
RT "Cloning of cDNA encoding human placental protein 12 (PP12): binding
RT protein for IGF I and somatomedin.";
RL Nucleic Acids Res. 16:8711-8711(1988).
RN [4]
RN SEQUENCE FROM N.A.
RC TISSUE=Decidua;
RX MEDLINE=88312985; PubMed=2457513;
RA Julkunen M., Koistinen R., Aalto-Setälä K., Seppälä M., Janne O.A.,
RA Kontula K.;
RT "Primary structure of human insulin-like growth factor-binding
RT protein/placental protein 12 and tissue-specific expression of its
RT mRNA.";
RL FEBS Lett. 236:295-302(1988).
RN [5]
RN SEQUENCE FROM N.A.
RX MEDLINE=88334540; PubMed=2458522;
RA Lee Y.-L., Hintz R.L., James P.M., Lee P.D.K., Shively J.E.,
RA Powell D.R.;
RT "Insulin-like growth factor (IGF) binding protein complementary
RT deoxyribonucleic acid from human HEP G2 hepatoma cells: predicted
RT protein sequence suggests an IGF binding domain different from those
RT of the IGF-I and IGF-II receptors.";
RL Mol. Endocrinol. 2:404-411(1988).
RN [6]
RN SEQUENCE FROM N.A.
RX MEDLINE=8930502; PubMed=2474129;
RA Cubbage M.L., Suwanichkul A., Powell D.R.;
RT "Structure of the human chromosomal gene for the 25 kilodalton
RT insulin-like growth factor binding protein.";
RL Mol. Endocrinol. 3:846-851(1989).
RN [7]
RN SEQUENCE FROM N.A.
RX MEDLINE=89087480; PubMed=2849945;
RA Brinkman A., Groffen C.A., Kortleve D.J., Drop S.L.S.;
RT "Organization of the gene encoding the insulin-like growth factor
RT binding protein IBP-1.";
RL Biochem. Biophys. Res. Commun. 157:898-907(1988).
RN [8]
RN SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=92217971; PubMed=1373120;
RA Ehrenborg E., Larsson C., Stern I., Janson M., Powell D.R.,

RA Luthman H.;
RT "Contiguous localization of the genes encoding human insulin-like
RT growth factor binding proteins 1 (IGBP1) and 3 (IGBP3) on chromosome
RT 7.";
RL Genomics 12:497-502(1992).
RN [9]
RN SEQUENCE OF 141-259 FROM N.A., AND SEQUENCE OF 26-259.
RC TISSUE=Amniotic fluid;
RX MEDLINE=89170723; PubMed=2466665;
RA Luthman H., Soederling-Barros J., Persson B., Engberg C., Stern I.,
RA Lake M., Franzen S.A., Israelsson M., Raden B., Lindgren B.,
RA Hjelqvist L., Enerbaeck S., Carlsson P., Bjursell G., Pova G.,
RA Hall K., Joernvall H.;
RT "Human insulin-like growth-factor-binding protein. Low-molecular-mass
RT form: protein sequence and cDNA cloning.";
RL Eur. J. Biochem. 180:259-265(1989).
RN [10]
RN SEQUENCE OF 26-53.
RX MEDLINE=89008261; PubMed=2971653;
RA Busby W.H. Jr., Klapper D.G., Clemmons D.R.;
RT "Purification of a 31,000-dalton insulin-like growth factor binding
RT protein from human amniotic fluid. Isolation of two forms with
RT different biologic actions.";
RL J. Biol. Chem. 263:14203-14210(1988).
RN [11]
RN MUTAGENESIS.
RX MEDLINE=92070504; PubMed=1959616;
RA Brinkman A., Kortlirve D.J., Schuller A.G.P., Zwarthoff E.C.,
RA Drop S.L.S.;
RT "Increase of beta-actin mRNA upon hypotonic perfusion of perfused rat
RT liver.";
RL FEBS Lett. 292:264-268(1991).
RN [12]
RN PHOSPHORYLATION SITES, AND PARTIAL SEQUENCE.
RX MEDLINE=93123224; PubMed=7678248;
RA Jones J.I., Busby W.H. Jr., Wright G., Smith C.E., Kinack N.M.,
RA Clemmons D.R.;
RT "Identification of the sites of phosphorylation in insulin-like
RT growth factor binding protein-1. Regulation of its affinity by
RT phosphorylation of serine 101.";
RL J. Biol. Chem. 268:1125-1131(1993).
RN [13]
RN DISULFIDE BONDS.
RX MEDLINE=99262603; PubMed=10329650;
RA Neumann G.M., Bach L.A.;
RT "The N-terminal disulfide linkages of human insulin-like growth
RT factor-binding protein-6 (hIGFBP-6) and hIGFBP-1 are different as
RT determined by mass spectrometry.";
RL J. Biol. Chem. 274:14587-14594(1999).
CC -!- FUNCTION: IGF-binding proteins prolong the half-life of the IGFs
CC and have been shown to either inhibit or stimulate the growth
CC promoting effects of the IGFs on cell culture. They alter the
CC interaction of IGFs with their cell surface receptors.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- PTM: Phosphorylated; probably by casein kinase II. Alters the
CC affinity of the protein for IGFs.
CC -!- MISCELLANEOUS: Binds equally well IGF-I and IGF-II.
CC -!- SIMILARITY: Contains 1 IGFBP domain.
CC -!- SIMILARITY: Contains 1 thyroglobulin type-1 domain.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to frameshifts
CC in positions 55 and 71.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; Y00856; CAA68770.1; -;
DR EMBL; M20841; AAA52540.1; ALT_FRAME.
DR EMBL; X12385; CAA30942.1; -;

DR EMBL: X13405; CAA31771.1; -
 DR EMBL: M31145; AAA52542.1; -
 DR EMBL: M59316; AAA52783.1; -
 DR EMBL: M23595; AAA52785.1; -
 DR EMBL: M23592; AAA52785.1; JOINED.
 DR EMBL: M23593; AAA52785.1; JOINED.
 DR EMBL: M23594; AAA52785.1; JOINED.
 DR EMBL: M23594; AAA52785.1; JOINED.
 DR EMBL: M74587; AAA52784.1; -
 DR EMBL: X15002; CAA33110.1; -
 DR PIR: A31867; IOHUL.
 DR HSP: P24593; LBOE.
 DR TRANSFAC: T00400; -
 DR Slena-2DPAGE; P08933; -
 DR Genew; HGNC:5469; IGFBP1.
 DR MIM; 146730; -
 DR GO: GO:0005615; C:extracellular space; TAS.
 DR GO: GO:0005520; F:insulin-like growth factor binding activity; TAS.
 DR GO: GO:0007165; P:signal transduction; TAS.
 DR InterPro: IPR000867; Ins_lgro_fac.pr.
 DR Pfam: PF00219; IGFBP; 1.
 DR Pfam: PF00086; thyroglobulin_1.
 DR SMART: SM00121; IB; 1.
 DR SMART: SM00211; TY; 1.
 DR PROSITE; PS00222; IGF_BINDING; 1.
 DR PROSITE; PS00484; THYROGLOBULIN_1; 1.
 KW Growth factor binding; Signal; Phosphorylation; Polymorphism.
 FT SIGNAL 1 25
 FT CHAIN 26 259 INSULIN-LIKE GROWTH FACTOR BINDING
 FT DOMAIN 202 251 PROTEIN 1.
 FT SITE 246 248 THYROGLOBULIN TYPE I.
 FT DISULFID 71 84 CELL ATTACHMENT SITE.
 FT DISULFID 78 104
 FT DISULFID 176 206
 FT MOD_RES 126 126 PHOSPHORYLATION (MAJOR).
 FT MOD_RES 144 144 PHOSPHORYLATION.
 FT MOD_RES 194 194 PHOSPHORYLATION.
 FT VARIANT 183 183 V -> I (IN dbSNP:1065782).
 FT VARIANT 253 253 I -> M (IN dbSNP:4619).
 FT VARIANT 253 253 /FTID-VAR_003821.
 FT CONFLICT 213 213 H -> Q (IN REF. 2).
 FT SEQUENCE 259 AA; 27903 MW; 8AA75AF7DC966012 CRC64;

Query Match 7.7%; Score 79; DB 1; Length 259;
 Best Local Similarity 28.3%; Pred. No. 2.6;
 Matches 26; Conservative 12; Mismatches 26; Indels 28; Gaps 4;

QY 12 RIVLNCAYAQTRGK-----GCOFIS--QTGRDKNOVEGFQIVSTAAQTFLATCIN 62
 DB 182 RVVESLAKAQTSGSEISKEYLPNCNKNGFHSROCTSMDEA----- 225
 QY 63 GVCWTYVHGAGTRTASPKR---GPVTQMYTNV 91
 DB 226 GLWCVCYVWNGKRIPGSPFIRGDPNCQIYENV 257

RESULT 14
 ID PAAD_PSEAE STANDARD; PRT; 209 AA.
 AC Q9HX08;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable aromatic acid decarboxylase (EC 4.1.1.-).
 GN PA4019.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 SEQUENCE FROM N.A.

RC STRAIN-ATCC 15692 / PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Huinagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goutry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT *Complete genome sequence of Pseudomonas aeruginosa PA01, an
 RT opportunistic pathogen.*;
 RL Nature 406:959-964(2000).
 CC Nature 406:959-964(2000).
 CC -!- SIMILARITY: BELONGS TO THE POLYPRENYL P-HYDROXYBENZOATE /
 CC PHENYLACRYLIC ACID DECARBOXYLASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE004818; AAG07406.1; -
 DR PIR: H83144; H83144.
 DR InterPro: IPR003382; Flavoprotein.
 DR Pfam: PF02441; Flavoprotein; 1.
 KW Hypothetical protein; Lyase; Decarboxylase; Complete proteome.
 SQ SEQUENCE 209 AA; 22367 MW; 01FD081CC495D3F6 CRC64;
 Query Match 7.7%; Score 78.5; DB 1; Length 209;
 Best Local Similarity 26.5%; Pred. No. 2.3;
 Matches 50; Conservative 16; Mismatches 56; Indels 67; Gaps 11;
 QY 41 QVEGEVO-IVSTAAQTFLATCINGVCVTVYHGAGTRTASPKGP----- 83
 DB 29 QEEREVHFLSKAAQLVNAI-----ETDVALPAKPAQAPLTYCGAAAG 74
 QY 84 VIQMYTNVDKLVGWPAQGSRLTP-----CTCGSSDL-----YLVTRHADVIPV 129
 DB 75 QIRVFGQND-----WMAPPASGSSAPNAWICPCTGSLSAVATGACNNLIERAADVALK 129
 QY 130 RRRGDSRGLSLSPR--PIS-----YLGSSGGPFLCPCAGHVCIFRAAVCTRGVAKVD 181
 DB 130 ER----RPLVLPREAPFSSIHLENMLKLSNIGAVILPA--APGFYHQ---POSVEDLVD 180
 QY 182 FIPVESLET 190
 DB 181 FVVARILNT 189
 RESULT 15
 ID ENV_FLVGL STANDARD; PRT; 642 AA.
 AC P08359;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ENV polyprotein precursor (Coat polyprotein) [Contains: Knob protein
 DE GP70; Spike protein P15E].
 GN ENV.
 OS Feline leukemia virus (strain A/Glasgow-1).
 OC Viruses; Retroid viruses; Retroviridae; Gammaretrovirus.
 OX NCBI_TaxID=11769;
 RN [1]
 SEQUENCE FROM N.A.
 RX MEDLINE=86200439; PubMed=3009890;
 RA Stewart M.A., Warnock M., Wheeler A., Wilkie N., Mullins J.I.,
 RA Onions D.E., Nell J.C.;
 RT *Nucleotide sequences of a feline leukemia virus subgroup A envelope
 RT gene and long terminal repeat and evidence for the recombinational
 RT origin of subgroup B viruses.*;
 RL J. Virol. 58:825-834(1986).
 CC -----

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 19:00:22 : Search time 37.2105 seconds
(without alignments)
1352.314 Million cell updates/sec

Title: US-09-965-594-12

Perfect score: 1021
Sequence: 1 MKKKGSVVIGRIYNGAYA.....VAKAVDFIPVESLETTMRSP 195

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_23:*

- 1: sp-archaea:*
- 2: sp-bacteria:*
- 3: sp-fungi:*
- 4: sp-human:*
- 5: sp-invertebrate:*
- 6: sp-mammal:*
- 7: sp-mmc:*
- 8: sp-organelle:*
- 9: sp-phage:*
- 10: sp-plant:*
- 11: sp-rodent:*
- 12: sp-virus:*
- 13: sp-vertebrate:*
- 14: sp-unclassified:*
- 15: sp-virus:*
- 16: sp-bacteriaph:*
- 17: sp-archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	908.5	89.0	4040	12 Q91FH8	Q91fh8 mucosal dis
2	894.5	87.6	2436	12 Q81756	Q81756 hepatitis c
3	894.5	87.6	3011	12 Q91FE5	Q91fe5 hepatitis c
4	893.5	87.5	3011	12 Q03463	Q03463 hepatitis c
5	891	87.3	181	12 Q91RR8	Q91rr8 hepatitis c
6	891	87.3	181	12 Q91RT5	Q91rt5 hepatitis c
7	889.5	87.1	3011	12 Q36579	Q36579 hepatitis c
8	889	87.1	181	12 Q91RR5	Q91rr5 hepatitis c
9	888	87.0	181	12 Q91RR3	Q91rr3 hepatitis c
10	888	87.0	181	12 Q91RS1	Q91rs1 hepatitis c
11	888	87.0	181	12 Q91RQ8	Q91rq8 hepatitis c
12	888	87.0	181	12 Q91RT1	Q91rt1 hepatitis c
13	886	86.8	181	12 Q91RR6	Q91rr6 hepatitis c
14	886	86.8	181	12 Q91RS9	Q91rs9 hepatitis c
15	885.5	86.7	3011	12 Q91RS8	Q91rs8 hepatitis c
16	885	86.7	181	12 Q91RR2	Q91rr2 hepatitis c

17	885	86.7	181	12 Q91RS3	Q91rs3 hepatitis c
18	884	86.6	181	12 Q91RT4	Q91rt4 hepatitis c
19	884	86.6	181	12 Q91RS8	Q91rs8 hepatitis c
20	884	86.6	181	12 Q91RT3	Q91rt3 hepatitis c
21	884	86.6	181	12 Q91RS5	Q91rs5 hepatitis c
22	884	86.6	181	12 Q91RS7	Q91rs7 hepatitis c
23	884	86.6	181	12 Q91RT0	Q91rt0 hepatitis c
24	882.5	86.4	3011	12 Q91RT6	Q91rt6 hepatitis c
25	882.5	86.4	3011	12 Q36608	Q36608 hepatitis c
26	882.5	86.4	3015	12 Q9PWX5	Q9pwx5 hepatitis c
27	882.5	86.4	3015	12 Q9PWU9	Q9pwx9 hepatitis c
28	882	86.4	181	12 Q91RS4	Q91rs4 hepatitis c
29	881	86.3	181	12 Q91RT6	Q91rt6 hepatitis c
30	880	86.2	181	12 Q91RT9	Q91rt9 hepatitis c
31	879	86.1	181	12 Q91RR4	Q91rr4 hepatitis c
32	879	86.1	181	12 Q91RR9	Q91rr9 hepatitis c
33	879	86.1	181	12 Q91RR0	Q91rr0 hepatitis c
34	877	85.9	181	12 Q91RR7	Q91rr7 hepatitis c
35	876.5	85.8	3011	12 Q36609	Q36609 hepatitis c
36	876	85.8	181	12 Q91RT2	Q91rt2 hepatitis c
37	876	85.8	181	12 Q91RR1	Q91rr1 hepatitis c
38	876	85.8	181	12 Q91RO9	Q91rq9 hepatitis c
39	876	85.8	181	12 Q91RS2	Q91rs2 hepatitis c
40	874	85.6	181	12 Q91RS6	Q91rs6 hepatitis c
41	873.5	85.6	3011	12 Q36610	Q36610 hepatitis c
42	873	85.5	181	12 Q91RT7	Q91rt7 hepatitis c
43	871	85.3	181	12 Q91RS0	Q91rs0 hepatitis c
44	871	85.3	181	12 Q91RT8	Q91rt8 hepatitis c
45	860.5	84.3	3011	12 Q81754	Q81754 hepatitis c

ALIGNMENTS

RESULT 1

ID	Q91FH8	PRELIMINARY:	PRT: 4040 AA.
AC	Q91FH8:		
DT	01-OCT-2000 (TReMBLrel. 15, Created)		
DT	01-OCT-2000 (TReMBLrel. 15, Last sequence update)		
DT	01-MAR-2003 (TReMBLrel. 23, Last annotation update)		
DE	Genome polyprotein.		
OS	Mucosal disease virus.		
OC	Viruses; ssRNA positive-strand viruses, no DNA stage: Flaviviridae;		
OX	Pestivirus.		
NCBI_TaxID	111099;		
RP	SEQUENCE FROM N.A.		
RP	MEDLINE-20323484; PubMed-10864644;		
RA	Lai V.C., Zhong W., Skelton A., Ingravallo P., Vassiliev V.,		
RA	Donis R.O., Hong Z., Lau J.Y.;		
RT	"Generation and characterization of a hepatitis C virus NS3 protease-		
RT	dependent bovine viral diarrhea virus.";		
RL	J. Virol. 74:6339-6347(2000).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RA	Lai V.C.H., Hong Z.;		
RL	Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: AF268278; AAF82566.1; -		
DR	HSSP: P26663; 1JXP.		
DR	MEROPS: S31.001; -		
DR	InterPro: IPR000280; CDvir_endptsep80.		
DR	InterPro: IPR001410; HCV_NS3.		
DR	InterPro: IPR004109; HCV_RdRP.		
DR	InterPro: IPR001666; HCV_RdRP.		
DR	InterPro: IPR001650; Helicase_C.		
DR	InterPro: IPR001005; Myb_DNA_Binding.		
DR	InterPro: IPR001568; RNase_T2.		
DR	InterPro: IPR007095; RNA_pol_DS_PS.		
DR	InterPro: IPR007094; RNA_pol_PSVir.		
DR	Pfam: PF02907; HCV_NS3; 1.		
DR	Pfam: PF00271; Helicase_C; 1.		
DR	Pfam: PF00998; Viral_RdRP; 1.		

```

DR PRINTS: PR00729; CDVENDOPINASE.
DR SMART: SM00487; DEXDc; 1.
DR SMART: SM00490; HELICc; 1.
DR PROSITE: PS00037; MYB_1; 1.
DR PROSITE: P50507; RORP_POSITIVE; 1.
DR PROSITE: P50521; RORP_VIRAL; 1.
DR PROSITE: PS00331; RNASE_T2; 1.
KW ATP-binding; Helicase; Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase.
SQ SEQUENCE 4040 AA; 453073 MW; ADE87791D055B9DC CRC64;

Query Match      89.0%; Score 908.5; DB 12; Length 4040;
Best Local Similarity 90.8%; Pred. No. 9e-84; Indels 5; Gaps 1
Matches 177; Conservative 6; Mismatches 7;

QY 5 GSWIVGRIVLNG-----AYAQOTRGEBCGOETISQTGRDNKNVEGEVQIVSTAATFLAT 59
   ||||| ||||| : | ||||| : ||||| ||||| ||||| ||||| ||||| |||||
Db 10 GSWIVGRIVLNLSGGSGITACAAQOTRGLLCKKITSLTGRDNKNVEGEVQIVSTATOTFLAT 69
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 60 CLINGVCWTVYHGAGTGRTIASPGKPIQMYTNVDKDLVGMPAPOGSRSLTPCTCGSSDDLVL 119
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 70 CLINGVCWTVYHGAGTGRTIASPGKPIQMYTNVDQDLVGMPAPOGSKSLTPCTCGSSDDLYI; 129
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 120 VTRHADVIPRRRGDSRGSLLSPRIPSYLKSGSGGPLLCFAGHAGVGFRAAVCTRGVAKA 179
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 130 VTRHANVIPRRRGDSRGSLLSPRIPSYLKSGSGGPLLCFAGHAGVGLFRAAVCTRGVAKA 189
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 180 VDFIPVESLETHMS 194
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 190 VDFIPVENLETTRS 204
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 2
Q81756 PRELIMINARY; PRT; 2436 AA.
ID Q81756
AC Q81756;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
CC NCBI_Taxid=11103;
RX [1]
RN Choo Q.-L., Richman K., Han J.;
RA "The nucleotide sequence of the Hepatitis C viral genome.";
RL Submitted (MAY-1990) to the EMBL/GenBank/DDBJ databases.
DR EMBL; H32084; AAA45677.1; .
DR HSSP; P27958; IAIY.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RDRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR007095; RNA_pol_PS.
DR InterPro; IPR007094; RNA_pol_Psvir.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00998; Viral_RdRp; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDc; 1.
DR PROSITE; PS0507; RDRP_POSITIVE; 1.

```



```
QY 172 CTRGVAKAVDFIPVESLETMRSP 195
Db 1185 CTRGVAKAVDFIPVESLETMRSP 1208

RESULT 5
Q91RR8 PRELIMINARY; PRT; 181 AA.
AC Q91RR8;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pt.1Y;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and Response to IFN of the NS3 Protease Gene from
RT Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF369235; AAK54560.1; -.
DR InterPro: IPR004109; HCV_NS3.
DR Pfam: PF02907; HCV_NS3; 1.
KW Protease.
FT NON_TER 1 1
FT NON_TER 181 181
SQ SEQUENCE 181 AA; 19130 MW; 85091869299B7C35 CRC64;

Query Match 87.3%; Score 891; DB 12; Length 181;
Best Local Similarity 96.6%; Pred. No. 1.le-83;
Matches 171; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 18 AYAAQTGEGCGQETISQTRDKNOVEGEVQIVSTAQTFLATCINGVCWTVYHGAGTRTI 77
Db 5 AYAAQTGELGCIITSLTGRDKNOVEGEVQIVSTAQTFLATCINGVCWTVYHGAGTRTI 64
QY 78 ASPKGPVIOMYTNVDKDLVGPAPQGSRLTPTCGSSDLYLVTRHADVIPVRRGDSRG 137
Db 65 ASPKGPVIOMYTNVDKDLVGPAPQGSRLTPTCGSSDLYLVTRHADVIPVRRGDSRG 124
QY 138 SLLSPRPISYLKSGSGPLLCAGHAGVIFRAAVCTRGVAKAVDFIPVESLETMRIS 194
Db 125 SLLSPRPISYLKSGSGPLLCAGHAGVIFRAAVCTRGVAKAVDFIPVENLETMRIS 181

RESULT 7
Q36579 PRELIMINARY; PRT; 3011 AA.
AC Q36579;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H7;
RX MEDLINE=97373636; PubMed=9228008;
RA Kolykhalov A.A., Agapov E.V., Blight K.J., Mihalik K., Feinstone S.M.,
RA Rice C.M.;
RT "Transmission of hepatitis C by intrahepatic inoculation with
RT transcribed RNA.";
RL Science 277:570-574(1997).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC EMBL; AF009606; AAB66324.1; -.
DR HSP; P27958; 1HEI.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRp.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR007095; RNA_pol_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01338; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RdRp; 1.
DR ProDom; PD186062; HCV_NS1; 1.
```



```

Matches 170; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 18 AYAAQTREGGCGCQETSOTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRTI 77
    ||||| || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5 AYAAQTRGLLGCITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRTI 64
QY 78 ASPKGPVIQMTYNTVDKDLVGPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRGDSRG 137
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 65 ASPKGPVIQMTYNTVDKDLVGPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRGDSRG 124
QY 138 SLLSPRPISYLGSSGGPILCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETMRS 194
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 125 SLLSPRPISYLGSSGGPILCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETMRS 181

RESULT 11
Q91RQ8 ID Q91RQ8 PRELIMINARY; PRT: 181 AA.
AC Q91RQ8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Pt.52;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
    Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF369245; AAK54570.1; -.
DR InterPro: IPR004109; HCV_NS3.
DR Pfam: PF02907; HCV_NS3; 1.
KW Protease.
FT NON_TER 1 1
FT NON_TER 181 181
SQ SEQUENCE 181 AA; 19144 MW; C0C91F1E2EB0B32 CRC64;

Query Match 87.0%; Score 888; DB 12; Length 181;
Best Local Similarity 96.0%; Pred. No. 2.3e-83;
Matches 170; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 18 AYAAQTREGGCGCQETSOTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRTI 77
    ||||| || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5 AYAAQTRGLLGCITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRTI 64
QY 78 ASPKGPVIQMTYNTVDKDLVGPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRGDSRG 137
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 65 ASPKGPVIQMTYNTVDKDLVGPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRGDSRG 124
QY 138 SLLSPRPISYLGSSGGPILCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETMRS 194
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 125 SLLSPRPISYLGSSGGPILCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETMRS 181

RESULT 12
Q91RT1 ID Q91RT1 PRELIMINARY; PRT: 181 AA.
AC Q91RT1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Pt.52;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
    Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF369245; AAK54570.1; -.
DR InterPro: IPR004109; HCV_NS3.
DR Pfam: PF02907; HCV_NS3; 1.
KW Protease.
FT NON_TER 1 1
FT NON_TER 181 181
SQ SEQUENCE 181 AA; 19144 MW; C0C91F1E2EB0B32 CRC64;

Query Match 87.0%; Score 888; DB 12; Length 181;
Best Local Similarity 96.0%; Pred. No. 2.3e-83;
Matches 170; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 18 AYAAQTREGGCGCQETSOTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRTI 77
    ||||| || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5 AYAAQTRGLLGCITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRTI 64
QY 78 ASPKGPVIQMTYNTVDKDLVGPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRGDSRG 137
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 65 ASPKGPVIQMTYNTVDKDLVGPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRGDSRG 124
QY 138 SLLSPRPISYLGSSGGPILCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETMRS 194
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 125 SLLSPRPISYLGSSGGPILCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETMRS 181

```

```

RC STRAIN-Pt.161;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
    Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF369222; AAK54547.1; -.
DR InterPro: IPR004109; HCV_NS3.
DR Pfam: PF02907; HCV_NS3; 1.
KW Protease.
FT NON_TER 1 1
FT NON_TER 181 181
SQ SEQUENCE 181 AA; 19114 MW; ABB90B5B3ABA4E26 CRC64;

Query Match 87.0%; Score 888; DB 12; Length 181;
Best Local Similarity 96.0%; Pred. No. 2.3e-83;
Matches 170; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 18 AYAAQTREGGCGCQETSOTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRTI 77
    ||||| || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5 AYAAQTRGLLGCITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRTI 64
QY 78 ASPKGPVIQMTYNTVDKDLVGPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRGDSRG 137
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 65 ASPKGPVIQMTYNTVDKDLVGPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRGDSRG 124
QY 138 SLLSPRPISYLGSSGGPILCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETMRS 194
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 125 SLLSPRPISYLGSSGGPILCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETMRS 181

RESULT 13
Q91RR6 ID Q91RR6 PRELIMINARY; PRT: 181 AA.
AC Q91RR6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Pt.3T;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
    Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF369237; AAK54562.1; -.
DR InterPro: IPR004109; HCV_NS3.
DR Pfam: PF02907; HCV_NS3; 1.
KW Protease.
FT NON_TER 1 1
FT NON_TER 181 181
SQ SEQUENCE 181 AA; 19101 MW; 614ADA8B0F33CCAF CRC64;

Query Match 86.8%; Score 886; DB 12; Length 181;
Best Local Similarity 95.5%; Pred. No. 3.7e-83;
Matches 169; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 18 AYAAQTREGGCGCQETSOTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRTI 77
    ||||| || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5 AYAAQTRGLLGCITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRTI 64
QY 78 ASPKGPVIQMTYNTVDKDLVGPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRGDSRG 137
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 65 ASPKGPVIQMTYNTVDKDLVGPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRGDSRG 124
QY 138 SLLSPRPISYLGSSGGPILCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETMRS 194
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 125 SLLSPRPISYLGSSGGPILCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETMRS 181

```

```
RESULT 14
Q91RS9          PRELIMINARY;      PRT:   181 AA.
AC Q91RS9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pt.174;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF369224; AAK54549.1; -.
DR InterPro; IPR004109; HCV_NS3.
DR Pfam; PF02907; HCV_NS3; 1.
KW Protease.
FT NON_TER      1      181
FT NON_TER      181      181
SQ SEQUENCE 181 AA; 19131 MW; 8BD7FC2769DBD635 CRC64;

Query Match      86.8%; Score 886; DB 12; Length 181;
Best Local Similarity 96.0%; Pred. No. 3.7e-83;
Matches 170; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 18 AVAQOTRGEGGCOETSGTGRDNQVGEVQIVSTAAQ:FLATCINGCVWTVYHGAGTRTI 77
DB 5 AVAQOTRGILGCIITSLTRDKRQDQVGEVQIVSTAAQ:FLATCINGCVWTVYHGAGTRTI 64

QY 78 ASPKGPVIOMYNVDKDLVGMWPAQGSRSRLTCTCGSSDLXLYVTRHADVIPVRRRGDSRG 137
DB 65 ASPKGPVIOMYNVDKDLVGMWPAQGSRSRLTCTCGSSDLXLYVTRHADVIPVRRRGDSRG :24

QY 138 SLLSPRPISYLGSSGGPLLCAGHAGVIFRAAVCFIRGAKAVDFIPVESLETTMRS 194
DB 125 SLLSPRPISYLGSSGGPLLCAGHAGVIFRAAVCFIRGAKAVDFIPVENLETTMRS 181

RESULT 15
Q91ELS8          PRELIMINARY;      PRT:   3011 AA.
AC Q91ELS8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=colonel;
RA Desai S.M., Devare S., Yamaguchi J.;
RT "Hepatitis C Virus.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AF290978; AAG02099.1; -.
DR HSP; P27958; IHE1.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
```

```
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRp.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00998; Viral_RdRp; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXdc; 1.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
DR PROSITE; PS05057; RDRP_POSITIVE; 1.
DR PROSITE; PS05021; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA; 327107 MW; A6BECF5A3B3EE13F CRC64;

Query Match      86.7%; Score 885.5; DB 12; Length 3011;
Best Local Similarity 84.3%; Pred. No. 1.4e-81;
Matches 172; Conservative 10; Mismatches 11; Indels 11; Gaps 2;

QY 3 KKGSVWIVG---RIVLNG-----AVAQOTRGEGGCOETSGTGRDNQVGEVQIVST 51
DB 1005 RGQEILGPADGWMVSGWRLAPITAYAAQOTRGLLGCIITSLTRDNQVGEVQIVST 1064

QY 52 AAQTFLATCINGCVWTVYHGAGTRTIIASPKGPVIQMTYINVDKDLVGMWPAQGSRSLTPTCT 111
DB 1065 ATQTFLATCINGCVWTVYHGAGTRTIIASPKGPVIQMTYINVDKDLVGMWPAQGSRSLTPTCT 1124

QY 112 CGSSDLXLYVTRHADVIPVRRRGDSRGSLSPRPISYLGSSGGPLLCAGHAGVIFRAAV 171
DB 1125 CGSSDLXLYVTRHADVIPVRRRGDSRGSLSPRPISYLGSSGGPLLCAGHAGVIFRAAV 1184

QY 172 CTRGVAKAVDFIPVESLETTMRSP 195
DB 1185 CTRGVAKAVDFIPVENLETTMRSP 1208

Search completed: August 30, 2003, 19:18:18
Job time : 38.2105 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run On: August 30, 2003, 19:18:33 ; Search time 2534.57 seconds
(without alignments)
3147.423 Million cell updates/sec

Title: US-09-965-594-12
Perfect score: 1021
Sequence: 1 MKKGSVIVIGRIVLNGAYA.....VAKAVDFIPVESLETTMRSP 195

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Zgapop 6.0 , Zgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 scqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-O=/cgn2.1/USPTO_spool/US09965594/runat_29082003.151919.28310/app_query.fasta.1.2872
-DB=GenEmbl -OPMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS-human40.cdi -LIST=45
-DOCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTPMT=pto -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09965594 -CGN_1.1.14686_@runat_29082003.151919.28310 -NCPU=6 -ICPU=3
-NO_MMAP -LARGQUERY -NEG.SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOC
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -YGAPOP=6
-YGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_on:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pt:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_on:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_to:*
27: em_sts:*
28: em_un:*

29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pin:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Query Length	DB ID	Description
1	928.5	90.9	12734	6	AR179057 Sequence
2	908.5	89.0	12734	14	AF268278
3	894.5	87.6	5360	6	AR118686 Sequence
4	894.5	87.6	5360	6	I06434
5	894.5	87.6	5360	6	I09328 Sequence 8
6	894.5	87.6	5360	6	AR118692 Sequence
7	894.5	87.6	6785	6	I06440
8	894.5	87.6	6785	6	I09329 Sequence 10
9	894.5	87.6	7310	6	AR118696 Sequence
10	894.5	87.6	7310	6	I09331
11	894.5	87.6	7310	14	HPCPOLYP
12	894.5	87.6	8316	6	AR118703 Sequence
13	894.5	87.6	8987	6	AR118728 Sequence
14	894.5	87.6	9185	6	AR118722 Sequence
15	894.5	87.6	9185	6	AR118723 Sequence
16	894.5	87.6	9185	6	BD091382
17	894.5	87.6	9185	6	I08294
18	894.5	87.6	9379	6	AR166930 Sequence 1
19	894.5	87.6	9379	6	AR301300 Sequence
20	894.5	87.6	9401	6	AR176483 Sequence
21	894.5	87.6	9401	6	BD080334
22	894.5	87.6	9401	6	E66593
23	894.5	87.6	9401	6	I71894
24	894.5	87.6	9401	6	I81885
25	894.5	87.6	9401	14	HPCPLYPRE
26	894.5	87.6	9609	12	AF387805
27	894.5	87.6	9609	12	AF387808
28	894.5	87.6	9618	14	AF271632
29	894.5	87.6	9646	12	AF387806
30	894.5	87.6	9693	12	AF387807
31	894	87.6	2058	6	AX395309
32	894	87.6	2058	6	AX454818
33	893.5	87.5	9502	6	E08263
34	893.5	87.5	9502	6	E08264
35	893.5	87.5	9502	14	HPCHCJ1
36	892	87.4	1932	6	AR127809
37	892	87.4	1932	6	BD081910
38	892	87.4	8157	6	AR127810
39	892	87.4	8157	6	BD081911
40	891.5	87.3	1998	6	AR145264
41	891.5	87.3	9424	14	AF511948
42	891	87.3	543	14	AF369218
43	891	87.3	543	14	AF369235
44	890	87.2	2061	6	AX441176
45	890	87.2	2061	6	AX467113

ALIGNMENTS

RESULT 1

AR179057	AR179057	Sequence 1	12734 bp	DNA	linear	PAT 20-APR-2002
LOCUS	AF268278.1	from patent US 6326137.				
DEFINITION	AF268278.1	GI:20220612				
ACCESSION	AF268278.1	GI:20220612				
VERSION	AF268278.1	GI:20220612				
KEYWORDS	Unknown.					
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	1 (bases 1 to 12734)					
AUTHORS	Hong, Z., Lai, V.C.H. and Lau, J.Y.N.					
TITLE	Hepatitis C virus, protease-dependent chimeric pestivirus					
JOURNAL	Patent: US 6326137-A 1 04-DEC-2001;					
FEATURES	Location/Qualifiers					
source	1..12734					
BASE COUNT	4032 a 2604 c 3295 g 2803 t					
ORIGIN	/organism="unknown"					
Alignment Scores:						
Pred. No.:	3,06e-66	Length:	12734			
Score:	928.50	Matches:	180			
Percent Similarity:	94.87%	Conservative:	5			
Best Local Similarity:	92.31%	Mismatches:	5			
Query Match:	90.94%	Indels:	5			
DB:	6	Gaps:	1			
US-09-965-594-12 (1-195) x AR179057 (1-12734)						
Qy	5	GlySerValValIleValGlyArgGluValLeuAsnGly	-----AlaTyr 19			
Db	413	GGTAGTGTGTTATTTGTTAGTATGTTTATCTGGTAGTGGTAGTACGCGGTAC	472			
Qy	20	AlaGlnGlnThrArgGlyGluGluCysGlnClnuThrSerGlnThrGlyArgAspLys	39			
Db	473	GCCACAGACAGAGAGGCGCTCTAGGCTGTAAGATCACCAGTCTGACTGCGCGGGACAAA	532			
Qy	40	AsnGlnValGluGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAlaThr	59			
Db	533	AACCAAGTGAGGCTGAGTCCAGATCGTGCAACTGCTACCAACACCTTCCTTGGCAACG	592			
Qy	60	CysIleAsnGlyValCysTsrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer	79			
Db	593	TGCATCAATGGGGTATGCTGGACTGTCTACACCGGGCGGACAGGACCATCGCATCA	652			
Qy	80	ProLysGlyProValIleGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrPro	99			
Db	653	CCCAAGGCTCCTGTCATCCAGATGATATACCAATGTGGACCAAGACCTTGTGGCTGGCC	712			
Qy	100	AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu	119			
Db	713	GCCTCTCAAGGTTCCCGCTCATTGACACCTGACCTGCGGCTCCCTCGGACCTTTACCTG	772			
Qy	120	ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu	139			
Db	773	GTTACGAGGACCGCGACGTCATCCCGTGGCGGCGGAGGTATAGCAGGGTAGCCTG	832			
Qy	140	LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro	159			
Db	833	CTTTGCGCCCGGCCCATTTCTACCTACCAAAAAGGCTCCTCGGGGGTCCGCTGTGTGCCCC	892			
Qy	160	AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla	179			
Db	893	CGGGACACCGCGTGGCGCTATTGAGGGCCCGGTGTGCACCCGTGAGTGGCCAAAGCG	952			
Qy	180	ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 194				
Db	953	GTGACCTTTATCCTGTGGAGAACCTTAGACAAACCATGAGATCC 997				
RESULT 2						
LOCUS	AF268278	12734 bp	RNA	linear	VRL 12-JUL-2000	
DEFINITION	Pestivirus type 1, complete genome.					

3'UTR
BASE COUNT 4030 a 2608 c 3293 g 2802 t 1 others
ORIGIN
VALGYYGYQALSKRRVPMITDIYTIEDORLEDTHLOYAPNAIKTGTCTELKELAS
GDVFKIMGAI5DYAAGLEFVKQAEKIKTAPLEKNEAKRGVQVRFIDSLNKKKE
IIRYCLMGHTAIYKSTAAIRLGHETAFATLVIKW:AFGGSVSDBVKCAADVLYVYV
MKNPSFPCDSOTQEGRFVASF1SALATVYTKWYHN:SKVDEPALAYLPVATSA
LMKFTPTRLSSVILSTIYKTYLSIRKGSXSDG:ILCTCISAAFEILSONPVSGVSM
LGVGAIAAHNAIESSEOKRTLLMKVYVKNFLDOAATDELKYENPEKIIIMALFEAVQII
GNPLUJLHLYGKVVYIKGWEAKELSERAGNLFILIMPEAFELLMDSQGRKNSGN
YIIDLIGLHGKQINRGLKMYLWAPAPFSCDWTSPDSERIRLDNTYLRVETRCGY
EMKAFKNGGKLTVEESGPELCRNRPGRPVNYRVTKYODNLRREIKPVAKLEGQVE
HYKGVTAIDYKSGKMLATDKWEVEHGVITRLAKRVYGVGNGAYLGDPPNHRALV
PRDCATITKNTVOFLPKMKKCAFTYDLTISNLTJLJELVHNNLFEKEIPTATVYTWL
AYTFVNEVDGILKPVLSERVIPDPVDINLOPEVOVOTSEVGIILIGRELLMTGVTP
VLEKVEPDASONSQVYIGLDEGNYPGGIOHTLITEEIHNRDARPEFTIMILGSNSIS
NRAKTARNINYTNDMPREIRDLMAAGRMVVALRDVDPSELSEMYDFKTFIDREALE
ALSJGQPKVKTVEAVENLIEQKQDEIPNMFASDDPVFLEVALKNDKYVLVDGVE
VKDQAKAGATDTRLIKVEGSRTYAMKLSWFIQASNKQMSLTPLELLRCPPAT
KSNKGHMASADOLAGQWNEPLGCGVHLCTIPARVKIHPYEAYLKLDFIEERSEKPR
VKDVIREHNNWILKIRFOGLNLTNNKMLNPKLSEQLDREGRRNRNHOIGTINS
AGIRLEKLPVIRAOTDTRTFHEAIRDKIDKSENRONPHELHKLLEIFHTIAOPTLKH
YGEVTEQLEAGINRKGAGLEKKNIGVEDESEKHLVEQLVRDLKAGRIKYETAI
PKNEKRDVSDQAGDLVVEKRPVIOYPPAKTRIAITKVNYNMVQVPIVGYEGK
TPLFNIPOKVRKFNDSNPEVAVSFDTKAMDTQVTSKDLOIIGFIOKYKKKWHKFI
DTJDMTEFVPIITADGEVYIRNGORSGCOPDTSAGNSMLNVLTMVAFCESTGVYK
SENRVARIVCGDGLFTEKGLGKFPANKQWILHENGKPKQITEGKMKVATRFED
IEFCSHTPVPVMSDSSHMAGRTAVILSKMATRSDSGERGTTAYEKAVAFSFL
MYSNNPLVRRIICLLVLSQOPETDPSKHATYYKGDPIGAYKDVIGRNLSELRGTGFEK
LANLNLSTIGTICWTKHTSKRIIOQVAIGKEGNLWVADRI:ISSKTHGLYIPDKGF
TIOGKHVEQLQRLRTETNPMVGVTGTERYKLGPTVNL:LLIRLKLMTAVGSS*
12509, 12734
BASE COUNT 4030 a 2608 c 3293 g 2802 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 1,35e-64 Length: 12734
Score: 908.50 Matches: 177
Percent Similarity: 93.8% Conservative: 6
Best Local Similarity: 90.77% Mismatches: 7
Query Match: 88.98% Indels: 5
DB: 14 Gaps: 1
US-09-965-594-12 (1-195) x AF268278 (1-12734)
QY 5 GlySerValValIleValGlyArgIleValLeuAAsnGly-----AlaTyr 19
Dbb 413 GGTAGTGTGTATTGTGTAGAGATTGTTTATCTGCTAGTGGPAGTATCAGCGCGTGC 472
QY 20 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGlnThrSerGlnThrGlyArgAspLys 39
Dbb 473 GCCCAGCAGCAGAGAGGCTCTCAGGGTGTAGATCACCAGTCTGACTGGCGGGGACAAA 532
QY 40 AsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAlaThr 59
Dbb 533 AACCAAGTGGAGGTGAGGTCCAGATCGTGCTCACTGCCAACCTCTCTGCGCAACG 592
QY 60 CysIleAAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 79
Dbb 593 TGCATCAATGGGGTATGCTGCACTCTCTACACCGGGCGGGAACGAGCACCATCCATCA 652
QY 80 ProLysGlyProValIleGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpPro 99
Dbb 653 CCCAAGGGTCCCTCATCCAGATGATACCAATGTGGACCAAGACTTGTGGGTGGGCC 712
QY 100 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 119
Dbb 713 GCTCTCAAGTTCCTCGCTCATGTACACCTGCACCTCGCGGTCTCTCGGACTTTACCTG 772
QY 120 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 139
Dbb 773 GTTACGAGCAGCCCAACGCTCATCCCGTGGCGCGCGAGGTGATAGCAGGGGTAGCCTG 832
QY 140 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 159
Dbb 833 CTTTCGCCCGGCCCATTTCTACCTAAAGGCTCTCTGCGGGTCCCGTGTGTGCGCC 892

Qy 160 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 179
Dbb 893 CGCGGACACGCCGTGGCGCTTATTACAGGCGCGGTGTGCACCGGTGAGTGGCCAGGCG 952
Qy 180 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 194
Dbb 953 GTGACTTTTCCCTGTGGAGAACCTAGACACACACGAGATCC 997
RESULT 3
AR118686
LOCUS AR118686 5360 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 53 from patent US 6150087.
ACCESSION AR118686
VERSION AR118686.1 GI:14100596
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 5360)
AUTHORS Chien,D.Y.
TITLE NABV diagnostics and vaccines
JOURNAL Patent: US 6150087-A 53 21-NOV-2000;
FEATURES Location/Qualifiers
source 1..5360
/organism="unknown"
BASE COUNT 1060 a 1623 c 1532 g 1145 t
ORIGIN
Alignment Scores:
Pred. No.: 7.42e-64 Length: 5360
Score: 894.50 Matches: 175
Percent Similarity: 89.71% Conservative: 8
Best Local Similarity: 85.78% Mismatches: 10
Query Match: 87.61% Indels: 11
DB: 6 Gaps: 2
US-09-965-594-12 (1-195) x AR118686 (1-5360)
QY 3 LysLysGlySerValValIleValGly-----ArgIleValLeuAAsnGly----- 17
Dbb 867 CGCAGGGCGGGAGATGACTGCTCGGCCACCCGATGGAATGCTTCCAGGGGGTGGAGG 926
QY 18 -----AlaTyrAlaGlnGlnThrArgGlyGluGlyCysGlnGlu 31
Dbb 927 TTGCTGGCGCCCATCACGGCGTACGCCAGCAGACAGAGGCGCTCTTAGGGTGCTAATC 986
QY 32 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 51
Dbb 987 ACCAGCCTAAGTGGCGGGACAAAACCAAGTGGAGGGTGAGGTCAGATTGTGCTCAACT 1046
QY 52 AlaAlaGlnThrPheLeuAlaThrCysIleAAsnGlyValCysTrpThrValTyrHisGly 71
Dbb 1047 GCTGCCCAACCTTCTCGGCAACGTGATCAATGGGGTGTGCTGCACTGTCTACACGGG 1106
QY 72 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 91
Dbb 1107 GCCGGAAGGAGGACCATCGCTCACCCNAGGGTCTCTGTCATCAGATGTATACCAATGTA 1166
QY 92 AspLysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThr 111
Dbb 1167 GACCAAGACCTTGTGGGTGGCGGCTCCGCAAGGTAGCGGCTCATTTGACACCCCTGCACT 1226
QY 112 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 131
Dbb 1227 TCGGGCTCTCGGACCTTTACCTGGTCACGAGGACCGCGATGTCATTCGCGCGCGG 1286
QY 132 ArgGlyAspSerArgGlySerLeuSerProArgProIleSerTyrLeuLysGlySer 151
Dbb 1287 CGGGGTGATACAGGGGAGCCTGCTGTGCGCGCGGCCCATTTCTACTTGAAGGCTCC 1346
QY 152 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 171

```
Db 1347 TCGGGGGTCCGCTGTGTGCCCCCGGGGACGCCGTGGGCATATTTAGGGCCGGGTG 1406
QY 172 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 191
Db 1407 TGCACCCGTGGAGTGGCTAAGGGGTGGACTTTATCCCTGTGGAGAACCCTAGAGACAACC 1466
QY 192 MetArgSerPro 195
Db 1467 ATGAGTCCCG 1478

RESULT 4
LOCUS 106434 5360 bp DNA linear PAT 02-DEC-1994
DEFINITION Sequence 48 from Patent EP 0318216.
ACCESSION 106434
VERSION 106434.1 GI:590311
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 5360)
AUTHORS Houghton,M., Choo,Q.-L. and Kuo,G.
TITLE Nanbv diagnostics and vaccines
JOURNAL Patent: EP 0318216-A1 48 31-MAY-1989;
FEATURES
source
LOCATION/Qualifiers
BASE COUNT 1061 a 1623 c 1533 g 1143 t
ORIGIN

Alignment Scores:
Pred. No.: 7.42e-64 Length: 5360
Score: 894.50 Matches: 175
Percent Similarity: 89.71% Conservative: 8
Best Local Similarity: 85.78% Mismatches: 10
Query Match: 87.61% Indels: 11
DB: 6 Gaps: 2

US-09-965-594-12 (1-195) x 106434 (1-5360)
QY 3 LysLysGlySerValValIleValGly-----ArgIleValLeuAsnGly----- 17
Db 867 CGCAGGGCCGGAGATACCTGCTCGGCCACGCCGATGTAATGCTCTCCAAGGGTGGAGG 926
QY 18 -----AlaTyrAlaGlnGlnThrArgGlyGluGluGlnGlu 31
Db 927 TTGCTGGCGCCCATCACGGGTACGCCACGACAGCAAGGGGCTCTAGGGTGCAATC 986
QY 32 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 51
Db 987 ACCAGCTTAACCTGCGCGGACAAACCAAGTGGAGGGTGAGGTCCAGATTGTGCAACT 1046
QY 52 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGly 71
Db 1047 GCTGCCCAACCTTCCTGGCAACGTGCATCAATGGGGTGTGCTGGACTGCTACCAAGGG 1106
QY 72 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 91
Db 1107 GCCGGAACGAGGACCATCGGCTCACCAAGGGTCTCTCATCCAGATGTATACCAATGTA 1166
QY 92 AspLysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThr 111
Db 1167 GACCAAGACCTTGTGGGCTGGCCGCTCCGCAAGGTAGCCGCTCATTTGACACCCCTGCAC 1226
QY 112 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 131
Db 1227 TGGCGCTCCTCGGACCTTTACCTGGTCAGAGGACGCCGATGTCATTCCTCGCGCGG 1286
QY 132 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 151
Db 1287 CGGGGTGATAGCAGGGGACGCTGCTGCCCGCGGCCCATTTCTACTTTGAAAGGCTCC 1346
QY 152 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 171
```

```
Db 1347 TCGGGGGTCCGCTGTGTGCCCCCGGGGACGCCGTGGGCATATTTAGGGCCGGGTG 1406
QY 172 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 191
Db 1407 TGCACCCGTGGAGTGGCTAAGGGGTGGACTTTATCCCTGTGGAGAACCCTAGAGACAACC 1466
QY 192 MetArgSerPro 195
Db 1467 ATGAGTCCCG 1478

RESULT 5
LOCUS 109328 5360 bp DNA linear PAT 02-DEC-1994
DEFINITION Sequence 8 from Patent WO 8904669.
ACCESSION 109328
VERSION 109328.1 GI:587963
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 5360)
AUTHORS Houghton,M., Choo,Q.-K. and Kuo,G.
JOURNAL Patent: WO 8904669-A 8 01-JUN-1989;
FEATURES
source
LOCATION/Qualifiers
BASE COUNT 1061 a 1623 c 1533 g 1143 t
ORIGIN

Alignment Scores:
Pred. No.: 7.42e-64 Length: 5360
Score: 894.50 Matches: 175
Percent Similarity: 89.71% Conservative: 8
Best Local Similarity: 85.78% Mismatches: 10
Query Match: 87.61% Indels: 11
DB: 6 Gaps: 2

US-09-965-594-12 (1-195) x 109328 (1-5360)
QY 3 LysLysGlySerValValIleValGly-----ArgIleValLeuAsnGly----- 17
Db 867 CGCAGGGCCGGAGATACCTGCTCGGCCACGCCGATGTAATGCTCTCCAAGGGTGGAGG 926
QY 18 -----AlaTyrAlaGlnGlnThrArgGlyGluGluGlnGlu 31
Db 927 TTGCTGGCGCCCATCACGGGTACGCCACGACAGCAAGGGGCTCTAGGGTGCAATC 986
QY 32 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 51
Db 987 ACCAGCTTAACCTGCGCGGACAAACCAAGTGGAGGGTGAGGTCCAGATTGTGCAACT 1046
QY 52 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGly 71
Db 1047 GCTGCCCAACCTTCCTGGCAACGTGCATCAATGGGGTGTGCTGGACTGCTACCAAGGG 1106
QY 72 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 91
Db 1107 GCCGGAACGAGGACCATCGGCTCACCAAGGGTCTCTCATCCAGATGTATACCAATGTA 1166
QY 92 AspLysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThr 111
Db 1167 GACCAAGACCTTGTGGGCTGGCCGCTCCGCAAGGTAGCCGCTCATTTGACACCCCTGCAC 1226
QY 112 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 131
Db 1227 TGGCGCTCCTCGGACCTTTACCTGGTCAGAGGACGCCGATGTCATTCCTCGCGCGG 1286
QY 132 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 151
Db 1287 CGGGGTGATAGCAGGGGACGCTGCTGCCCGCGGCCCATTTCTACTTTGAAAGGCTCC 1346
QY 152 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 171
```



```
|||||
Db 1347 TCGGGGGTCCGCTGTGTGCCCCCGGGGACGCGGTGGGCATATTATAGGGCCGCGGTG 1406
Qy 172 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThr 191
Db 1407 TGCACCGGTGGAGTGGCTAGGCGGTGACATTATCCCTGTGGAGAACCTAGAGACAACC 1466
Qy 192 MetArgSerPro 195
Db 1467 ATGAGGTCCCG 1478

RESULT 6
LOCUS AR118692 6785 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 65 from patent US 6150087.
ACCESSION AR118692
VERSION AR118692.1 GI:14100602
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 6785)
AUTHORS Chien,D.Y.
TITLE NANBV diagnostics and vaccines
JOURNAL Patent: US 6150087-A 65 21-NOV-2000;
FEATURES
    source
        1. .6785
        /organism="unknown"
BASE COUNT 1392 a 2050 c 1914 g 1429 t
ORIGIN

Alignment Scores:
Pred. No.: 9,59e-64 Length: 6785
Score: 894.50 Matches: 175
Percent Similarity: 89.71% Conservative: 8
Best Local Similarity: 85.78% Mismatches: 10
Query Match: 87.61% Indels: 11
DB: Gaps: 2

US-09-965-594-12 (1-195) x AR118692 (1-6785)
Qy 3 LysLysGlySerValValIleValGly-----ArgIleValLeuAsnGly----- 17
Db 1140 CGCAGGGCGGGAGATACTGCTCGGGCCAGCGCATGGTATGCTCCAAAGGGGTGGAG 1199
Qy 18 -----AlaTyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlu 31
Db 1200 TTGCTGGCGGCCATACAGGGGTACGCCAGCAGACAAGGGGCTCTAGGGTGCAATC 1259
Qy 32 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGluValGlnIleValSerThr 51
Db 1260 ACCAGCCTAACTCGCGCGGACAAAACCAAGTGGAGGTGAGTCCAGATTGTGTCACT 1319
Qy 52 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGly 71
Db 1320 GCTGCCCAAACTTCTCGCAAGCTGCATCAATGGGGTGTCTGGACTGTCTACACGGG 1379
Qy 72 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 91
Db 1380 GCCGAAGCAGGACCACGCGCTCACCAAGGGTCTGTATCCAGATGATATACCAATGTA 1439
Qy 92 AspLysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThr 111
Db 1440 GACCAAGACCTTGTGGGCTGGCGGCTCCGCAAGGTAGCGGCTCATTTGACACCTGCAC 1499
Qy 112 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 131
Db 1500 TCGGGCTCTCGGACCTTTACCTGGTCACGAGGACGCCGATGTCAATCCCGTGGCGCGG 1559
Qy 132 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 151
Db 1560 CGGGGTGATAGCAGGGGACGCTGCTGTCGCCCGGCCCATTTCTTACTTGAAGGCTCC 1619
```

```
Qy 152 SerClyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaIleVal 171
Db 1620 TCGGGGGTCCGCTGTGTGCCCCCGGGGACGCGGTGGGCATATTATAGGGCCGCGGTG 1679
Qy 172 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThr 191
Db 1680 TGCACCGGTGGAGTGGCTAGGCGGTGACATTATCCCTGTGGAGAACCTAGAGACAACC 1739
Qy 192 MetArgSerPro 195
Db 1740 ATGAGGTCCCG 1751

RESULT 7
LOCUS I06440 6785 bp DNA linear PAT 02-DEC-1994
DEFINITION Sequence 54 from Patent EP 0318216.
ACCESSION I06440
VERSION I06440.1 GI:590312
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 6785)
AUTHORS Houghton,M., Choo,O.-L. and Kuo,G.
TITLE Nanbv diagnostics and vaccines
JOURNAL Patent: EP 0318216-A1 54 31-MAY-1989;
FEATURES
    Location/Qualifiers
        source
            1. .6785
            /organism="unknown"
BASE COUNT 1392 a 2050 c 1914 g 1429 t
ORIGIN

Alignment Scores:
Pred. No.: 9,59e-64 Length: 6785
Score: 894.50 Matches: 175
Percent Similarity: 89.71% Conservative: 8
Best Local Similarity: 85.78% Mismatches: 10
Query Match: 87.61% Indels: 11
DB: Gaps: 2

US-09-965-594-12 (1-195) x I06440 (1-6785)
Qy 3 LysLysGlySerValValIleValGly-----ArgIleValLeuAsnGly----- 17
Db 1140 CGCAGGGCGGGAGATACTGCTCGGGCCAGCGCATGGTATGCTCCAAAGGGGTGGAG 1199
Qy 18 -----AlaTyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlu 31
Db 1200 TTGCTGGCGGCCATACAGGGGTACGCCAGCAGACAAGGGGCTCTAGGGTGCAATC 1259
Qy 32 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGluValGlnIleValSerThr 51
Db 1260 ACCAGCCTAACTCGCGCGGACAAAACCAAGTGGAGGTGAGTCCAGATTGTGTCACT 1319
Qy 52 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGly 71
Db 1320 GCTGCCCAAACTTCTCGCAAGCTGCATCAATGGGGTGTCTGGACTGTCTACACGGG 1379
Qy 72 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 91
Db 1380 GCCGAAGCAGGACCACGCGCTCACCAAGGGTCTGTATCCAGATGATATACCAATGTA 1439
Qy 92 AspLysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThr 111
Db 1440 GACCAAGACCTTGTGGGCTGGCGGCTCCGCAAGGTAGCGGCTCATTTGACACCTGCAC 1499
Qy 112 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 131
Db 1500 TCGGGCTCTCGGACCTTTACCTGGTCACGAGGACGCCGATGTCAATCCCGTGGCGCGG 1559
Qy 132 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 151
Db 1560 CGGGGTGATAGCAGGGGACGCTGCTGTCGCCCGGCCCATTTCTTACTTGAAGGCTCC 1619
```

```
Qy 152 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 171
|||||
Db 1620 TCGGGGGTCCGGTGTGTGCCCCCGGGGACGGCGTGGCATATTATTAGGGCCGGGTG 1679

Qy 172 CysThrArgGlyValAlaLysAlaValAspPheIleProValGlySerLeuGluThrThr 191
|||||
Db 1680 TGCACCCGTGGAGTGGCTAAGCGGTGGACITTTATCCCTGTGGAGAACCTAGAGACAACC 1739

Qy 192 MetArgSerPro 195
|||||
Db 1740 ATGAGGTCCCGG 1751

RESULT 8
109329
LOCUS 109329 6785 bp DNA linear PAT 02-DEC-1994
DEFINITION Sequence 10 from Patent WO 8904669.
ACCESSION 109329
VERSION 109329.1 GI:587964
KEYWORDS
SOURCE Unknown.
ORGANISM
REFERENCE 1 (bases 1 to 6785)
AUTHORS Houghton,M., Choo,Q.-K. and Kuo,G.
JOURNAL Patent: WO 8904669-A 10 01-JUN-1989;
FEATURES
source 1. .6785
BASE COUNT 1392 a 2050 c 1914 g 1429 t
ORIGIN

Alignment Scores:
Pred. No.: 9.59e-04 Length: 6785
Score: 894.50 Matches: 175
Percent Similarity: 89.71% Conservative: 8
Best Local Similarity: 85.78% Mismatches: 10
Query Match: 87.61% Indels: 11
DB: 6 Gaps: 2

US-09-965-594-12 (1-195) x 109329 (1-6785)

Qy 3 LysLysGlySerValValIleValGly-----ArgIleValLeuAsnGly----- 17
|||||
Db 1140 CGCAGGGCGGGAGATACTGCTGGGCCAGCCGATGGATGGTCTCCAAGGGGTGGAGG 1199

Qy 18 -----AlaTyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlu 31
|||||
Db 1200 TTGCTGGCGCCCATCAGCGGTACGCCCGCAGACAGAGGGGCCCTCTTAGGGTGCATATC 1259

Qy 32 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 51
|||||
Db 1260 ACCAGCCTAAGTGGCGGGACAAAACCAAGTGGAGGGTGAAGTCCAGATTGTGCAACT 1319

Qy 52 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGly 71
|||||
Db 1320 GCTGCCCAACCTTCTTGGCAACGTGCATCAATGGGTGTGCTGGAGTGTCTACCAACGG 1379

Qy 72 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 91
|||||
Db 1380 GCCGGAACGAGGACCATCGGTCAACCAAGGGTCTGTCTCCAGATGTATACCAATGTA 1439

Qy 92 AspLysAspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThr 111
|||||
Db 1440 GACCAAGACCTTGTGGGTGGCGGCTCCGCAAGGTAGCGGCTCATTCACACCCCTGCACT 1499

Qy 112 CysGlySerSerAspLeuThrValThrArgHisAlaAspValIleProValArgArg 131
|||||
Db 1500 TCGCGCTCCTCGGACCTTACCTGGTCACGAGGACGCCGATGTCATTCCTCGCGCGG 1559

Qy 132 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 151
|||||
Db 1560 CGGGGTGATACAGGGGAGCGCTGCTGTGCGCCCGGCCCATTTCTTACTTGAAAGGCTCC 1619
```

```
Qy 152 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 171
|||||
Db 1620 TCGGGGGTCCGGTGTGTGCCCCCGGGGACGGCGTGGCATATTATTAGGGCCGGGTG 1679

Qy 172 CysThrArgGlyValAlaLysAlaValAspPheIleProValGlySerLeuGluThrThr 191
|||||
Db 1680 TGCACCCGTGGAGTGGCTAAGCGGTGGACITTTATCCCTGTGGAGAACCTAGAGACAACC 1739

Qy 192 MetArgSerPro 195
|||||
Db 1740 ATGAGGTCCCGG 1751

RESULT 9
AR118696
LOCUS AR118696 7310 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 74 from patent US 6150087.
ACCESSION AR118696
VERSION AR118696.1 GI:14100606
KEYWORDS
SOURCE Unknown.
ORGANISM
REFERENCE 1 (bases 1 to 7310)
AUTHORS Chien,D.Y.
TITLE NANBY diagnostics and vaccines
JOURNAL Patent: US 6150087-A 74 21-NOV-2000;
FEATURES
source 1. .7310
BASE COUNT 1495 a 2220 c 2056 g 1539 t
ORIGIN

Alignment Scores:
Pred. No.: 1.04e-63 Length: 7310
Score: 894.50 Matches: 175
Percent Similarity: 89.71% Conservative: 8
Best Local Similarity: 85.78% Mismatches: 10
Query Match: 87.61% Indels: 11
DB: 6 Gaps: 2

US-09-965-594-12 (1-195) x AR118696 (1-7310)

Qy 3 LysLysGlySerValValIleValGly-----ArgIleValLeuAsnGly----- 17
|||||
Db 1665 CGCAGGGCGGGAGATACTGCTGGGCCAGCCGATGGATGGTCTCCAAGGGGTGGAGG 1724

Qy 18 -----AlaTyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlu 31
|||||
Db 1725 TTGCTGGCGCCCATCAGCGGTACGCCCGCAGACAGAGGGGCCCTCTTAGGGTGCATATC 1784

Qy 32 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 51
|||||
Db 1785 ACCAGCCTAAGTGGCGGGACAAAACCAAGTGGAGGGTGAAGTCCAGATTGTGCAACT 1844

Qy 52 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGly 71
|||||
Db 1845 GCTGCCCAACCTTCTTGGCAACGTGCATCAATGGGTGTGCTGGAGTGTCTACCAACGG 1904

Qy 72 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 91
|||||
Db 1905 GCCGGAACGAGGACCATCGGTCAACCAAGGGTCTGTCTCCAGATGTATACCAATGTA 1964

Qy 92 AspLysAspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThr 111
|||||
Db 1965 GACCAAGACCTTGTGGGTGGCGGCTCCGCAAGGTAGCGGCTCATTCACACCCCTGCACT 2024

Qy 112 CysGlySerSerAspLeuThrValThrArgHisAlaAspValIleProValArgArg 131
|||||
Db 2025 TCGCGCTCCTCGGACCTTACCTGGTCACGAGGACGCCGATGTCATTCCTCGCGCGG 2084

Qy 132 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 151
|||||
```

```
Db 2085 CGGGGTGATAGCAGGGGCGAGCTGCTGTGCGCCCGGGCCCATTTCCCTACTTGAAGAGCTCC 2144
Qy 152 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 171
Db 2145 TCGGGGGTCCGCTGTGTCGCCCGGGGACGCGCTGGGCATATTTAGGCGCGGGTG 2204
Qy 172 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThr 191
Db 2205 TGCACCCGTGGAGTGGCTAAGCGGTGGACTTTATCCCTGTGGAGAACCTAGACACAACC 2264
Qy 192 MetArgSerPro 195
Db 2265 ATGAGGTCCCG 2276

RESULT 10
LOCUS I09331 7310 bp DNA linear PAT 02-DEC-1994
DEFINITION Sequence 15 from Patent WO 8904669.
ACCESSION I09331
VERSION I09331.1 GI:587966
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 7310)
AUTHORS Houghton, M., Choo, Q.-K. and Kuo, G.
JOURNAL Patent: WO 8904669-A 15 01-JUN-1989;
FEATURES
    Location/Qualifiers
    source
        /organism="unknown"
BASE COUNT 1495 a 2218 c 2058 g 1539 t
ORIGIN

Alignment Scores:
Pred. No.: 1 04e-63 Length: 7310
Score: 894.50 Matches: 175
Percent Similarity: 89.71% Conservative: 8
Best Local Similarity: 85.78% Mismatches: 10
Query Match: 87.61% Indels: 11
DB: 6 Gaps: 2

US-09-965-594-12 (1-195) x I09331 (1-7310)
Qy 3 LysLysGlySerValValIleValGly-----ArgIleValLeuAsnGly----- 17
Db 1665 CGCAGGCGCGGAGATAGTCTCGCGCGGCGGATGGAATGCTCCCAAGGGT*GGAGG 1724
Qy 18 -----AlaTyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlu 31
Db 1725 TTGCTGGCGCCCATCAGCGGTCAGCGCCAGCAGACAGGGGGCTCTCTAGGGTGCATAATC 1784
Qy 32 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGlnValGlnIleValSerThr 51
Db 1785 ACCAGCTTACTGCGCGCGGCAAAACCAAGTGGAGGGTGAAGTCCAGATTGTGTCNACT 1844
Qy 52 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGly 71
Db 1845 GCTGCCCAACCTTCTCGCAACGTGCATCAATGAGGTGTGCTGGAGTCTGTACACCGG 1904
Qy 72 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 91
Db 1905 GCGCGAAGCAGGACCATCGCGTCACCAAGGTCCTGTCTATCCAGATGATACCAATGTA 1964
Qy 92 AspLysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThr 111
Db 1965 GACCAAGACCTTGTGGGTGGCGCGCTCCGCAAGTAGGCGCTCATTGACACCTGCACCT 2024
Qy 112 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 131
Db 2025 TGGCGCTCTCGACCTTTACCTGGTGCAGGACGACGCGCATGTCAATCCCGTGCACCG 2084
Qy 132 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 151
```

```
Db 2085 CGGGGTGATAGCAGGGGCGAGCTGCTGTGCGCCCGGGCCCATTTCCCTACTTGAAGAGCTCC 2144
Qy 152 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 171
Db 2145 TCGGGGGTCCGCTGTGTCGCCCGGGGACGCGCTGGGCATATTTAGGCGCGGGTG 2204
Qy 172 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThr 191
Db 2205 TGCACCCGTGGAGTGGCTAAGCGGTGGACTTTATCCCTGTGGAGAACCTAGACACAACC 2264
Qy 192 MetArgSerPro 195
Db 2265 ATGAGGTCCCG 2276

RESULT 11
LOCUS HPCPOLYP 7310 bp ss-RNA linear VRL 02-AUG-1993
DEFINITION Hepatitis C virus polyprotein gene, partial cds.
ACCESSION M32084
VERSION M32084.1 GI:329875
KEYWORDS polyprotein.
SOURCE Hepatitis C virus
ORGANISM Hepatitis C virus
REFERENCE 1 (bases 1 to 7310)
AUTHORS Choo, Q.-L., Richman, K. and Han, J.
JOURNAL The nucleotide sequence of the Hepatitis C viral genome
COMMENT Unpublished (1990)
COMMENT Original source text: Hepatitis C virus, cDNA to viral RNA, clones K9-1 through 15e, isolated from chimpanzee (individual 910) blood plasma.
Draft entry and printed sequence for [1] kindly submitted by M.Houghton, 22-FEB-1990. Chiron Corporation, 4560 Horton Street, Emeryville CA 94608.
FEATURES
    Location/Qualifiers
    source
        1..7310
        /organism="Hepatitis C virus"
        /mol_type="genomic RNA"
        /db_xref="taxon:11103"
        <1..>7310
        /note="polyprotein"
        /codon_start=3
        /protein_id="AAA45677.1"
        /db_xref="GI:329876"
        /translation="GCPELRSCLPLDFQGWGPISYANGSGPDQRPYCMWHPKPC
        GIVPAKSVCGPVCTPSPVVGTTDRSGAPTSWGENDTDFVFNNTRPPLGNWFC
        TWMNSTGCTKVCAGAPCVIGGAGNNTLHCPTDCPKHPDATYSCSGSPWITPRCLVD
        YPRLWHYPCPTINYTIPIKIRMYGVGVHLEPAACNMTGRSCDLEDKRSFSLPLLT
        ITQOVLPCEFTTLPALSTGLIHQNLVDVQVLYGVGSSIASHAIKWEYVWLLFLLL
        ADARVCSCLMMMLLIQAALLENILVILNAASLATHGLVSLVFFCFAMTLKGKWP
        GAVTYFCMNPPLLLLLALPQRAYALDTEVAASCGVVLVGLMALTLSPPYKRYISM
        LWMYQFLTRVEAQLHVMIPPLNVGRGRDAVILLCAVHPTLVFDITKLLAVFGLPL
        ILOASLLKPVFVRVQGLRFLCALARKMIGHYVQVWVLIKLALTYVYVNHLPDRLD
        WAHNGRLDVAVEPVFVSQMETKLITMGADTAACGDIINGLPVSARGREILLPLRD
        GMSKGRLLAPITAYAAQTQKGLGITSITSLGDRKNQVEVEVIVSTAAQTFLATCI
        NGVCTVTHGAGTRTIASPGPVIOMYTNVDQDLGWPAQPSGSLTPTCTGSSDLVL
        TRHADVTPRRRGDSRGLSPRPISYLKSGSGGLPCPAGHAGVIFRAACTGTGVA
        KAVDFIPVENLETTMRSPVTDNSPPVPOFVAHLHAPTGSKSTKYPAAYAAOG
        KYVLVLPVAATIGKAYMSKANGIDPNIRGTITITGSPITYSTYGFGLDGXGS
        GARDIICDECHSDATSLIGTIVLQDAETAGARLWALATATPPSVVYVPHNIEE
        VALSTGEIPYKRAIPLVETKGRHLIFCHSKKDELAALVALGINAVATYTRGLD
        VSVPTSGDVVVAVATDAIMTGTGDFSDVSDCNTVOTVDFSLDPTFTLETITLPD
        AVSRTORGRTRGKPGIYREVAPGERSGDFSSVLCYECYDAGCANYELTPAETVVR
        LRAYMTPGLPVCODHLEFEGVFTGLTHIDAFLSQTQKSGENLPYLVAQATVCAR
        AQAPPSPDQMWKMLIRLKLPTLHGPTPLLYBLGAVONEITLTHPTVYINTCMSADLE
        VYTSWLVGVGLAALAAVCLSTGCVIVGVRLSGRPALIDPREDVYRFEDEMEES
        QHLPYIEGMMMLAEQFKKALGLQATASROAEVIAVQTNQKLETFMAKMMWNFIS
        GIQLAGLSLTPGNPATLMAFTAAVTSPLTSTQTLFNTLFGWVAAQALAAQCAATA
        FVGAGLCAAGIAGVGLKVLIDILAGYAGAGVAGALVAFKIMSGVSPSTEDVNLPLAI
        LSPGALVGVVCAALIRHRVGGEGAVOMMNRLLAFASRGNHVSPTHVPESDAARV
        TALSLSLVTLQRLRLHWWISSSECTPCGCSWLRDINDWICEVLSDFKTLWKLAKMFP
        LPIGPFVSCQRYGKGVWRVVDGIMHTRCHGAEITGVKNGMTMRIVGPRTCKRMWGLFQ
```

PINAVTTGCTPLPAPNTFALWRVSAEYVEIROVGFPHYVYVGMTDNLKPCQVPS
PEFFTELQVRLHREAPPCLLREEVFRVLGHEYPVGSQLPCEPEPDVAVLTSMLT
DPSHTAEAGRLRARGSPSVASSASOLSAPSLKATCTANHDSFDAELIGANLWR
QEMGNIITVESENKVLDSFDPLVAFEDEREISVPAEILLKRSRFAQALPVWAPRD
YNPLVETWKKDIEPPVPHGCPPLPPKPPPPVPRKRTVVLTSTLSTALAEATR
SFGSSSTGTDGNTITTCSEAPSPSCDAESYSSMPPLEGEPDPLSDGSWSTY
SSEANEDYVCCMSYSWTGALVTPCAAEOKLPINALNSLLRHINIIVYSTTSAC
QOKKVTFORLOVLDHYQDLVEKAAAKVKNLLSVEEACSLTPHSAKSKFYG
ANDRCHARKMALYHNSVKOLLIEDNVTPIDTTINAKNEVFCVQPEKGRKPARLIV
PDLVGRCKMALYDVTKLPLAVMCSYVGFQYSPQORVERFLVQAWKSKLTPMFPYFD
TRCFDSTVESDIRTEALYQCCDDIDPOARVAIKSLTERLYKYGGLPLTNRGNCGRYR
CRASGVLTSCCNTITCYIKARAARAAGLDQCTMLVCGDDLVVICESAGVEDAASL
RAFTAMTRYAPPQPEYDLELITSCSNVSAHNDGAKRVYITRTDPTTLAR
AAWETARHTPVSNSLGNITWFAPTLWARMILMTFFSVLIARDQLEQALDCBIYAGY
SIEPLDLPPIIOLR*

BASE COUNT 1495 a 2218 c 2058 g 1539 t
ORIGIN

Alignment Scores:
Pred. No.: 1,04e-63 Length: 7310
Score: 894.50 Matches: 175
Percent Similarity: 89.71% Conservative: 8
Best Local Similarity: 85.78% Mismatches: 10
Query Match: 87.61% Indels: 11
DB: 14 Gaps: 2

US-09-965-594-12 (1-195) x HPCPOLYP (1-7310)

QY 3 LysLysGlySerValValIleValGly-----ArgIleValLeuAsnGly----- 17
Db 1665 CGCAGGGCGGGAGATACTGCTCGGGCCAGCGATGGAAATGGTCTCCAAGGGGTGGAGG 1724
QY 18 -----AlaTyrAlaGlnGlnThrArgGlyGluGluGlyGlnGlu 31
Db 1725 TTGCTGGCGCCCATCAGCGGTACGCCACACAGAGGGGCTCTAGGTGCATAATC 1784
QY 32 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 51
Db 1785 ACCAGCTAATTCGCGGGACAAACCAAGTGGAGGGTGAGGTCCAGATTTGTCAACT 1844
QY 52 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGly 71
Db 1845 GCTGCCCAACCTTCTTGGCAAGTGCATCAATGGGGTGTGCTGGACTGTCTACCAAGGG 1904
QY 72 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 91
Db 1905 GCCGGAACGAGGACCATCGCGTCACCAAGGGTCTGTCTCATCCAGATGATACCAATGTA 1964
QY 92 AspLysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThr 111
Db 1965 GACCAAGACCTTTGGGGCTGGCGCGCTCCGCAAGGTAGCGGCTCATTTGACACCTGCAT 2024
QY 112 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 131
Db 2025 TGGCGTCTCGGACCTTTACCTTGGTCACAGGACGCCGATGTCATTCCTGCGCGCGG 2084
QY 132 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 151
Db 2085 CGGGGTGATAGCAGGGCAGCCTGCTGTGCGCGCGGCCCATTTCTACTTTGAAAGGCTCC 2144
QY 152 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 171
Db 2145 TCGGGGGTCCGCTGTGTGCGCGCGGGCAGCGCGTGGGCATATTATGAGGGCGCGGTG 2204
QY 172 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 191
Db 2205 TGCACCGGTGGGTGGGTAGGGCGGTGGACTTTATCCCTGTGGAGAACCTTAGAGACAAACC 2264
QY 192 MetArgSerPro 195
Db 2265 ATGAGGTCCCGG 2276

RESULT 12

AR118703
LOCUS AR118703 8316 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 88 from patent US 6150087.
ACCESSION AR118703
VERSION AR118703.1 GI:14100613
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 8316)
AUTHORS Chien,D.Y.
TITLE NAMV diagnostics and vaccines
JOURNAL Patent: US 6150087-A 88 21-NOV-2000;
FEATURES Location/Qualifiers
1..8316
Source /organism="unknown"

BASE COUNT 1671 a 2529 c 2345 g 1771 t
ORIGIN

Alignment Scores:
Pred. No.: 1.2e-63 Length: 8316
Score: 894.50 Matches: 175
Percent Similarity: 89.71% Conservative: 8
Best Local Similarity: 85.78% Mismatches: 10
Query Match: 87.61% Indels: 11
DB: 6 Gaps: 2

US-09-965-594-12 (1-195) x AR118703 (1-8316)

QY 3 LysLysGlySerValValIleValGly-----ArgIleValLeuAsnGly----- 17
Db 2671 CGCAGGGCGGGAGATACTGCTCGGGCCAGCGATGGAAATGGTCTCCAAGGGGTGGAGG 2730
QY 18 -----AlaTyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlu 31
Db 2731 TTGCTGGCGCCCATCAGCGGTACGCCACAGACAGAGGGGCTCTAGGTGCATAATC 2790
QY 32 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 51
Db 2791 ACCAGCTAATTCGCGGGACAAACCAAGTGGAGGGTGAGGTCCAGATTTGTCAACT 2850
QY 52 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGly 71
Db 2851 GCTGCCCAACCTTCTTGGCAAGTGCATCAATGGGGTGTGCTGGACTGTCTACCAAGGG 2910
QY 72 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 91
Db 2911 GCCGGAACGAGGACCATCGCGTCACCAAGGGTCTGTCTCATCCAGATGATACCAATGTA 2970
QY 92 AspLysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThr 111
Db 2971 GACCAAGACCTTTGGGGCTGGCGCGCTCCGCAAGGTAGCGGCTCATTTGACACCTGCAT 3030
QY 112 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 131
Db 3031 TCGGGTCTCGGACCTTTACCTTGGTCACAGGACGCCGATGTCATTCCTGCGCGCGG 3090
QY 132 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 151
Db 3091 CGGGGTGATAGCAGGGCAGCCTGCTGTGCGCGCGGCCCATTTCTACTTTGAAAGGCTCC 3150
QY 152 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 171
Db 3151 TCGGGGGTCCGCTGTGTGCGCGCGGGCAGCGCGTGGGCATATTATGAGGGCGCGGTG 3210
QY 172 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 191
Db 3211 TGCACCGGTGGGTGGGTAGGGCGGTGGACTTTATCCCTGTGGAGAACCTTAGAGACAAACC 3270
QY 192 MetArgSerPro 195
Db 3271 ATGAGGTCCCGG 3282

```
RESULT 13
ARI18728
LOCUS ARI18728 8987 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 137 from patent US 6150087.
ACCESSION ARI18728
VERSION ARI18728.1 GI:14100638
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 8987)
AUTHORS Chien,D.Y.
TITLE NANBV diagnostics and vaccines
JOURNAL Patent: US 6150087-A 137 21-NOV-2000;
FEATURES
    source
        1..8987
            /organism="unknown"
BASE COUNT 1807 a 2735 c 2547 g 1898 t
ORIGIN

Alignment Scores:
Pred. No.: 1..3e-63 Length: 8987
Score: 894.50 Matches: 175
Percent Similarity: 89.71% Conservative: 8
Best Local Similarity: 85.78% Mismatches: 10
Query Match: 87.61% Indels: 11
DB: Gaps: 2

US-09-965-594-12 (1-195) x ARI18728 (1-8987)
QY 3 LysLysGlySerValValIleValGly-----ArgIleValLeuAsnGly----- 17
Db 3013 CGCAGGCGCGGAGATGCTCGCGCCAGCGCATGGAATGCTCCAAAGGGGTGGAGG 3072
QY 18 -----AlaTyrAlaGlnGlnThrArgGlyGluGluCysGlnGlu 31
Db 3073 TTGCTGGCGCCCATCATCGGCGTACGCCACGACAGCAAGGGCCCTCTAGGTGCATATC 3132
QY 32 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGluValGlnIleValSerThr 51
Db 3133 ACCAGCCTAACTGCGCGGACAAAACCAAGTGGAGGTCAGTCCAGATTGTGTCAACT 3192
QY 52 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValIleHisGly 71
Db 3193 GCTGCCCAAACTTCCTGCCAACGTGCATCAATGGGCTGTGCTGGACTGTCTACCAACGG 3252
QY 72 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 91
Db 3253 GCCGGAACGAGGACCATCGCGTCACCCAAAGGTCCTGTATCCAGATGTATACCAATGTA 3312
QY 92 AspLysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThr 111
Db 3313 GACCAACACCTTGTGGCTGGCCGCTCCGCAAGTAGCGGTCAATGACACCTGCAC 3372
QY 112 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArg 131
Db 3373 TCGCGCTCTCGGACCTTACCTGGTCACGAGGACGCCGATGCTATCCCGTCGCGCGG 3432
QY 132 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 151
Db 3433 CGGGGTGATACGAGGCGACCTGCTGTGCGCCGCGCCATTCCTACTTTGAAGGCTCC 3492
QY 152 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 171
Db 3493 TCGGGGGTCCGCTGTTGTGCGCCGCGGCGACCGCTGGGCATATTAGGGCGCGCGTG 3552
QY 172 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 191
Db 3553 TGCACCGGTGGAGTGGCTAAGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACC 3612
QY 192 MetArgSerPro 195
Db 3613 ATGAGGTCCCGC 3624
```

```
RESULT 14
ARI18722
LOCUS ARI18722 9185 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 122 from patent US 6150087.
ACCESSION ARI18722
VERSION ARI18722.1 GI:14100632
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 9185)
AUTHORS Chien,D.Y.
TITLE NANBV diagnostics and vaccines
JOURNAL Patent: US 6150087-A 122 21-NOV-2000;
FEATURES
    source
        1..9185
            /organism="unknown"
BASE COUNT 1849 a 2790 c 2608 g 1938 t
ORIGIN

Alignment Scores:
Pred. No.: 1..33e-63 Length: 9185
Score: 894.50 Matches: 175
Percent Similarity: 89.71% Conservative: 8
Best Local Similarity: 85.78% Mismatches: 10
Query Match: 87.61% Indels: 11
DB: Gaps: 2

US-09-965-594-12 (1-195) x ARI18722 (1-9185)
QY 3 LysLysGlySerValValIleValGly-----ArgIleValLeuAsnGly----- 17
Db 3332 CGCAGGCGCGGAGATGCTCGCGCCAGCGCATGGAATGCTCCAAAGGGGTGGAGG 3391
QY 18 -----AlaTyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlu 31
Db 3392 TTGCTGGCGCCCATCATCGGCGTACGCCACGACAGCAAGGGCCCTCTAGGTGCATATC 3451
QY 32 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGluValGlnIleValSerThr 51
Db 3452 ACCAGCCTAACTGCGCGGACAAAACCAAGTGGAGGTCAGATTGTGTCAACT 3511
QY 52 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValIleHisGly 71
Db 3512 GCTGCCCAAACTTCCTGCCAACGTGCATCAATGGGCTGTGCTGGACTGTCTACCAACGG 3571
QY 72 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 91
Db 3572 GCCGGAACGAGGACCATCGCGTCACCCAAAGGTCCTGTATCCAGATGTATACCAATGTA 3631
QY 92 AspLysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThr 111
Db 3632 GACCAACACCTTGTGGCTGGCCGCTCCGCAAGTAGCGGTCAATGACACCTGCAC 3691
QY 112 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArg 131
Db 3692 TCGCGCTCTCGGACCTTACCTGGTCACGAGGACGCCGATGCTATCCCGTCGCGCGG 3751
QY 132 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 151
Db 3752 CGGGGTGATACGAGGCGACCTGCTGTGCGCCGCGCCATTCCTACTTTGAAGGCTCC 3811
QY 152 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 171
Db 3812 TCGGGGGTCCGCTGTTGTGCGCCGCGGCGACCGCTGGGCATATTTAGGGCGCGGTG 3871
QY 172 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 191
Db 3872 TGCACCGGTGGAGTGGCTAAGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACC 3931
QY 192 MetArgSerPro 195
Db 192 MetArgSerPro 195
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: August 30, 2003, 19:13:57 ; Search time 181.082 Seconds
(without alignments)
2905.924 Million cell updates/sec

Title: US-09-965-594-12

Perfect score: 1021

Sequence: 1 MKKGSVWIVGRVINGAYA.....VAKAVDFIPVESLETMRSP 195

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=rlp
-Q/cgn2_1/USPTO.spool/US09965594/runat_29082003_151918_28302/app_query.fasta_1.2872
-DB=N_Geneseq_19Jun03 -OFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=humad40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-NO_MMAPP -LARGE_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq_19Jun03:*
1: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA2002.DAT:*
25: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1021	100.0	588	21	AAA73329 Hepatitis C virus
2	998	97.7	594	21	AAA73330 Hepatitis C virus
3	982	96.2	588	21	AAA73328 Hepatitis C virus
4	981	96.1	594	21	AAA73331 Hepatitis C virus
5	959	93.9	594	21	AAA73335 Hepatitis C virus
6	951	93.1	594	21	AAA73332 Hepatitis C virus
7	939	92.0	594	21	AAA73333 Hepatitis C virus
8	929	91.0	594	21	AAA73334 Hepatitis C virus
9	928.5	90.9	12734	24	ABA95615 Chimeric BVDV/HCV
10	918.5	90.0	612	25	ABX15706 Anti-viral synthet
11	894.5	87.6	5300	10	AAH92097 Combined open read
12	894.5	87.6	5360	10	AAH90327 Hepatitis C virus
13	894.5	87.6	6905	10	AAH92103 Combined open read
14	894.5	87.6	7310	10	AAH92106 Combined open read
15	894.5	87.6	7310	10	AAH90336 Composite hepatiti
16	894.5	87.6	7310	16	AAH98221 Hepatitis C virus
17	894.5	87.6	8316	21	AAH75296 cDNA sequence comp
18	894.5	87.6	9133	20	AAZ07656 Nucleotide sequenc
19	894.5	87.6	9185	11	AAQ05956 Sense strand of th
20	894.5	87.6	9185	12	AAQ10566 Hepatitis C virus
21	894.5	87.6	9185	21	AAH75297 Sense strand of HC
22	894.5	87.6	9401	13	AAQ21744 Compiled HCV cDNA.
23	894.5	87.6	9401	17	AAH12710 Hepatitis C virus
24	894.5	87.6	9401	18	AAH99981 HCV polyprotein co
25	894.5	87.6	9401	19	AAV09989 Hepatitis C virus
26	894.5	87.6	9401	24	AAH35043 Hepatitis C virus
27	894	87.6	549	21	AAA70344 HCV-1 NS3/4a mutan
28	894	87.6	2058	24	AAH29795 Hepatitis C virus
29	894	87.6	2058	24	ABK15344 DNA encoding HCV-1
30	894	87.6	2058	25	ABK14410 Hepatitis C virus
31	893.5	87.4	9502	15	AAO74770 Hepatitis C virus
32	892	87.4	1933	20	AAH23258 HCV NS3 DNA. Hepa
33	892	87.4	8145	20	AAH23259 Plasmid pET-BS(+)/
34	891.5	87.3	1998	20	AAH80355 HCV NS4A-NS3 compl
35	891.5	87.3	9185	20	AAH26737 Nucleotide sequenc
36	891.5	87.3	9185	20	AAH00459 Hepatitis C virus
37	890.5	87.2	8316	11	AAQ05955 Hepatitis C virus
38	890	87.2	2061	24	AAH34500 Hepatitis C virus
39	890	87.2	2061	24	AAH31767 Hepatitis C virus
40	889.5	87.1	9646	19	AAV59361 Hepatitis C virus
41	889.5	87.1	9646	24	ABK87285 cDNA encoding hepa
42	889.5	87.1	12980	19	AAV59364 Hepatitis C virus
43	889.5	87.1	12980	24	ABK87286 Hepatitis C virus
44	889.5	87.1	16622	21	AAH36212 Nucleotide sequenc
45	888.5	87.0	1998	20	AAH80359 HCV NS4A-NS3 compl

ALIGNMENTS

RESULT 1	
AAA73329	
ID	AAA73329 standard; DNA; 588 BP.
XX	
AC	AAA73329;
XX	
XX	
DT	19-DEC-2000 (first entry)
XX	
DE	Hepatitis C virus NS4A-NS3 fusion protease coding sequence #2.
XX	
KW	Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW	liver failure; liver cancer; mutant; mutain; ds.
XX	
OS	Hepatitis C virus.
OS	Synthetic.
XX	
PH	Key Location/Qualifiers


```
XX SQ Sequence 594 BP; 103 A; 186 C; 156 G; 149 T; 0 other;
Alignment Scores:
Pred. No.: 5,64e-82 Length: 594
Score: 998.00 Matches: 193
Percent Similarity: 98.48% Conservativity: 1
Best Local Similarity: 97.97% Mismatches: 1
Query Match: 97.75% Indels: 2
DB: 21 Gaps: 1

US-09-965-594-12 (1-195) x AAA73330 (1-594)
QY 1 MetLysLysLysGlySerValValIleValGlyArgIleValLeuAsnGly-----Ala 18
DQ 1 ATGAAAAAAGAGTCCGTTGTTATCGTCGGCGGTATCAACCTGTCGGTGACACCGCT 60
QY 19 TyrAlaGlnGlnThrArgGlyGluGlyCysGlnGluThrSerGlnThrGlyArgasp 38
DQ 61 TACGCTCAGCAGCTCGAGGTGAGGAGGTGTCGAAGAACCTCCACACCGGTGCTGAC 120
QY 39 LysAsnGlnValGluGlyValGlnIleValSerThrAlaAlaGlnThrPheLeuAla 58
DQ 121 AAAACACAGTTGAAGTGAAGTTTCAGATCGTTTCCACCGCTGCAGACCTTCTGCGT 180
QY 59 ThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 78
DQ 181 ACCTGCTCAACGGTGTTCGTGGACCGTTTACCACGGTGTGTGACCGTACCACGCT 240
QY 79 SerProLysGlyProValIleGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp 98
DQ 241 TCCCGAAGGTCGGTTATCCAGATGTACACCAACGTTGACAAAGACCTGTTGTTGG 300
QY 99 ProAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 118
DQ 301 CGGCTCCGACGGTCCCGTTCCTGACCGCGTGCACCTGCGGTTCCTCCACCTGTAC 360
QY 119 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 138
DQ 361 CTGGTTACCGGTACGCTGACGTATCCCGGTTCGTCGTCGTCGTCGTCGTCGTCGTC 420
QY 139 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyCysProLeuLeuCys 158
DQ 421 CTGCTGTCCCGCGTCGATCTCCACCTGAAGGTTCCTCGGTGTCGTCGTCGTCGTCG 480
QY 159 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys 178
DQ 481 CCGGCTGTCGACGCTGTCGTCATCTCCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 540
QY 179 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 195
DQ 541 GCTGTTGACTTCATCCCGGTTGAATCCCTGGAACACCATCGCTTCCCGC 591

RESULT 3
AAA73328
ID AAA73328 standard; DNA; 588 BP.
AC
XX
XX
XX 19-DEC-2000 (first entry)
DE
XX Hepatitis C virus NS4A-NS3 fusion protease coding sequence #1.
XX Hepatitis C virus NS3 fusion protease; viral replication; chronic liver disease;
KW liver failure; liver cancer; ds.
XX
XX Hepatitis C virus.
OS Synthetic.
XX
XX Key Location/Qualifiers
FH 1..588
FT CDS
FT
FT /product= "NS3-NS4A fusion protein"
```

```

Db 361 ACCGCTACGCTGACGTTATCCCGTGTGCGTGGTGAATCCCGTGGTTCCTGCTG 420
QY 141 SerProArgProfileSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysProAla 160
DB 421 TCCCGCGGTGCGATCTCTACCTCAAGGTTCTCCCGTGTGCGTGTGCTGCGCGCT 480
QY 161 GlyHisAlaValGlyIlePheArgAlaValCysThrArgGlyValAlaLysAlaVal 180
DB 481 GGTCAAGCTGTGTGATCTTCCGTCGCTGCTTTGACACCGTGTGTGTTAAAGCTGT 540
QY 181 AspPheIleProValGluSerLeuGluThrThrMetArgSerPro 195
DB 541 GACTTCATCCCGGTGAATCCCTGGAACCAACCATGCGTTCCCGG 585

RESULT 4
ID AAA73331 standard; DNA; 594 BP.
XX
AC AAA73331;
DT 19-DEC-2000 (first entry)
XX
DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #4.
XX
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW liver failure; liver cancer; mutant; mutin; ds.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1..594
FT /tag= a
FT /product= "NS4A-NS3 fusion protein #4"
XX
WO200040707-A1.
XX
PD 13-JUL-2000.
XX
PF 06-JAN-2000; 2000MO-US00345.
XX
PR 08-JAN-1999; 99US-0115271.
XX
PA (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX
WPI; 2000-465976/40.
DR P-PSDB; AAB15222.
XX
PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
PT amino acid, useful for screening inhibitors that may treat hepatitis C
PT
XX
Claim 26; Fig 14; 66pp; English.
XX
XX
CC The present sequence is the coding sequence for a mutated version of a
CC fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
CC protease enzymes. These proteins are both essential for the replication
CC of the virus, acting to cleave its replicative proteins from the
CC polyprotein produced from the HCV genome. Inhibitors of the two proteins
CC should be effective as antiviral treatments of HCV infection. This is
CC useful as HCV can lead to chronic liver disease such as cirrhosis, liver
CC failure and liver cancer. The present invention concerns a number of NS3
CC mutants and NS3-NS4A fusion proteins which can be used to identify
CC inhibitors of this type, as well as enabling structural studies of the
CC protease and protease-inhibitor complexes. The protein produced from this
CC sequence contains the alpha-helix0-1 variant.
XX
SQ Sequence 594 BP; 105 A; 187 C; 155 G; 147 T; 0 other;
Alignment Scores:

```

```

Pred. No.: 1.99e-80 Length: 594
Score: 981.00 Matches: 190
Percent Similarity: 96.95% Conservative: 1
Best Local Similarity: 96.45% Mismatches: 4
Query Match: 96.08% Indels: 2
DB: 21 Gaps: 1

US-09-965-594-12 (1-195) x AAA73331 (1-594)
QY 1 MetLysLysLysGlySerValValIleValGlyArgGlyLeuValLeuAsnGly-----Ala 18
DB 1 ATGAAAAAAGGATCCCGTGTGTCGCGCGTATCACTGTCCGGTGACACCGCT 60
QY 19 TyrAlaGlnGlnThrArgGlyGluGlyCysGlnGlnThrSerGlnThrGlyArgGly 38
DB 61 TAGGCTCAGCAGACTCGAGGTGAGGAGGTGCCAAGAAACCTCCAGACCGTCTGCTG 120
QY 39 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAla 58
DB 121 AAAACCCAGGTTGAAGGTGAAGTTCAGATCGTTCCACCGCTACCCAGACCTTCTGCT 180
QY 59 ThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 78
DB 181 ACCTGATCAACGGTGTGTCGGACCGTTTACACGGTGTGTCGGTACCGGTACCACTGCT 240
QY 79 SerProLysGlyProValIleGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp 98
DB 241 TCCCCGAAAGGTCGCGTTACCCAGATGTACACCAAGTTGACAAAGACCTGTTGGTTGG 300
QY 99 ProAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 118
DB 301 CAGGCTCCGACGGGTTCCCGTTCCTCCGTCACCGGTGACCTGCGGTTCCTCCGACCTGAC 360
QY 119 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 138
DB 361 CTGGTTACCGCTCAGCGTCAGCTATATCCCGTTCGTCGTCGTCGTCGTCGTCGTCGTC 420
QY 139 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 158
DB 421 CTGCTGTCCCGCGTCCGATCTCTACCTGAAAGGTTCTCCGCGTGTCCGCTGCTGCTGC 480
QY 159 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys 178
DB 481 CCGGCTGGTCAAGCTGTGTAATCTTCCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 540
QY 179 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 195
DB 541 GCTGTTGACTTCATCCCGGTGAATCCCTGGAACCAACCATGCGTTCCCGG 591

RESULT 5
AAA73335
ID AAA73335 standard; DNA; 594 BP.
XX
AC AAA73335;
XX
DT 19-DEC-2000 (first entry)
XX
DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #8.
XX
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW liver failure; liver cancer; mutant; mutin; ds.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1..594
FT /tag= a
FT /product= "NS4A-NS3 fusion protein #8"
XX
WO200040707-A1.
XX
PD 13-JUL-2000.

```

Db	421	CTGCTGTCCTCCGCGCGCATCTCTACTGAAAGGTTCTCCGGTGGTCGCTGTCG 480
Qy	159	ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys 178
Db	481	CCGGCTGGTCACGCTGGTGTATCTTCGGTGTCTGCTTTGCACCGTGGTGTGCTAAA 540
Qy	179	AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 195
Db	541	GCTGTTGACTTCATCCCGGTTGAATCCCTGGAACCAACCATGCGTTCCCG 591
RESULT 6		
Id	AAAT73332	
XX	AAAT73332 standard; DNA; 594 BP.	
AC	AAAT73332;	
XX	19-DEC-2000 (first entry)	
DT		
XX	Hepatitis C virus NS4A-NS3 fusion protease coding sequence #5.	
DE		
XX	Hepatitis; NS3 protease; viral replication; chronic liver disease;	
KW	liver failure; liver cancer; mutant; mutein; ds.	
XX		
OS	Hepatitis C virus.	
OS	Synthetic.	
XX		
FH	Key Location/Qualifiers	
FT	CDS 1..594	
FT	/*tag= a	
FT	/product= "NS4A-NS3 fusion protein #5"	
XX		
XX	WO2000040707-A1.	
PN		
XX		
PD	13-JUL-2000.	
XX		
XX	06-JAN-2000; 2000WO-US00345.	
PF		
XX		
PR	08-JAN-1999; 99US-0115271.	
XX		
XX	(BRIM) BRISTOL-MYERS SQUIBB CO.	
PA		
XX	Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;	
PI		

XX WPI: 2000-465976/40.
 DR P-PSDB: AAB15223.
 XX
 XX
 PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 PT
 XX
 XX
 PS Claim 26; Fig 15; 66pp; English.
 XX
 CC The present sequence is the coding sequence for a mutated version of a
 CC fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
 CC protease enzymes. These proteins are both essential for the replication
 CC of the virus, acting to cleave its replicative proteins from the
 CC polyprotein produced from the HCV genome. Inhibitors of the two proteins
 CC should be effective as antiviral treatments of HCV infection. This is
 CC useful as HCV can lead to chronic liver disease such as cirrhosis, liver
 CC failure and liver cancer. The present invention concerns a number of NS3
 CC mutants and NS3-NS4A fusion proteins which can be used to identify
 CC inhibitors of this type, as well as enabling structural studies of the
 CC protease and protease:inhibitor complexes. The protein produced from this
 CC sequence contains the alpha-helix0-1 variant.
 XX
 SQ Sequence 594 BP; 105 A; 189 C; 153 G; 147 T; 0 other;

Alignment Scores:	1.07e-77	Length:	594
Pred. No.:	951.00	Matches:	187
Score:	95.43%	Conservative:	1
Percent Similarity:			

Best Local Similarity: 94.92% Mismatches: 7
Query Match: 93.14% Indels: 2
DB: 21 Gaps: 1

US-09-965-594-12 (1-195) x AAA73332 (1-594)

```
QY 1 MetLysLysGlySerValValIleValGlyArgIleValLeuAsnGly-----Ala 18
DB 1 ATGAAAAAAGGATCGGTTGTTATCGCGCGTATCAACCTGTCCGGTGACACCGCT 60
QY 19 TyrAlaGlnGlnThrArgGlyGluGlyCysGlnGlnThrSerGlnThrGlyArgAsp 38
DB 61 TACGCTCAGCAGACTCGAGGTGAGAGGGTTGCAAGAAACCTCCAGACGGTCTGAC 120
QY 39 LysAsnGlnValGluGlyValGlnIleValSerThrAlaAlaGlnThrPheLeuAla 58
DB 121 AAAAAACAGGTTGAAGTGAAGTTGATCGTTCCACCGGTACCCAGACCTTCTCGCT 180
QY 59 ThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrIleAla 78
DB 181 ACCTCCATCAACGGTGTCTGCGACCGTTTACCACGGTGTGTACCGGTACCATCGCT 240
QY 79 SerProLysGlyProValIleGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp 98
DB 241 TCCCGAAGGTCCCGTTTACCAGATGTACACCAACGTTGACAAAGACCTGGTTGGTGG 300
QY 99 ProAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 118
DB 301 CAGGCTCCGACGGTTCCTGTCACCGCGTTCGACCTGCGGTCTCTCCGACCTGTAC 360
QY 119 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 138
DB 361 CTGTTACCGCTCAGCTGACGTTATCCCGGTTCGTCGTCGTGACTCCGCTGGTTC 420
QY 139 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLysCys 158
DB 421 CTGCTGTCCCGCGTCCGATCTCTACCTGAAGGTTCCTCCGGTGGTCCGCTGTGTC 480
QY 159 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys 178
DB 481 CCGGCTGGTCAGCGTGTGGTATCTTCGTCGTCTGCTGTTCCACCGGTGGTGTCTAA 540
QY 179 AlaValAspPheIleProValGluSerLeuGlnThrThrMetArgSerPro 195
DB 541 GCTGTTGACTTCATCCCGGTTGAATCCCTCGAAACACCATCGCTTCCCG 591
```

RESULT 7

AAA73333

ID AAA73333 standard; DNA; 594 BP.

XX AC AAA73333;

XX 19-DEC-2000 (first entry)

XX Hepatitis C virus NS4A-NS3 fusion protease coding sequence #6.

XX Hepatitis; NS3 protease; viral replication; chronic liver disease;

XX liver failure; liver cancer; mutant; mutein; ds.

XX Hepatitis C virus.

XX Synthetic.

XX Key Location/Qualifiers

XX CDS 1..594

XX /*tag= a

XX /product= "NS4A-NS3 fusion protein #6"

XX MO2000040707-A1.

XX 13-JUL-2000.

XX 06-JAN-2000; 2000WO-US00345.

XX

PR 08-JAN-1999; 99US-0115271.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;

XX WPI; 2000-465976/40.

XX P-PSDB; AAB15224.

PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1 substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic amino acid, useful for screening inhibitors that may treat hepatitis C.

PS Claim 26; Fig 16; 66pp; English.

CC The present sequence is the coding sequence for a mutated version of a fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These proteins are both essential for the replication of the virus, acting to cleave its replicative proteins from the polyprotein produced from the HCV genome. Inhibitors of the two proteins should be effective as antiviral treatments of HCV infection. This is useful as HCV can lead to chronic liver disease such as cirrhosis, liver failure and liver cancer. The present invention concerns a number of NS3 mutants and NS3-NS4A fusion proteins which can be used to identify inhibitors of this type, as well as enabling structural studies of the protease and protease-inhibitor complexes. The protein produced from this sequence contains the alpha-helix0-7 variant.

SQ Sequence 594 BP; 104 A; 191 C; 152 G; 147 T; 0 other;

Alignment Scores:

Pred. No.: 1.32e-76 Length: 594
Score: 939.00 Matches: 184
Percent Similarity: 94.92% Conservative: 3
Best Local Similarity: 93.40% Mismatches: 8
Query Match: 91.97% Indels: 2
DB: 21 Gaps: 1

US-09-965-594-12 (1-195) x AAA73333 (1-594)

```
QY 1 MetLysLysGlySerValValIleValGlyArgIleValLeuAsnGly-----Ala 18
DB 1 ATGAAAAAAGGATCGGTTGTTATCGCGCGTATCAACCTGTCCGGTGACACCGCT 60
QY 19 TyrAlaGlnGlnThrArgGlyGluGlyCysGlnGlnThrSerGlnThrGlyArgAsp 38
DB 61 TACGCTCAGCAGACTCGAGGTGAGAGGGTTGCGAAGACCTCCACACCGGTCTGAC 120
QY 39 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAla 58
DB 121 AAAAAACAGGTTGAAGTGAAGTTGATCGTTTCCACCGGTACCCAGACCTTCTCGCT 180
QY 59 ThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrIleAla 78
DB 181 ACCTCCATCAACGGTGTCTGCGACCGTTTACCACGGTGTGTACCGGTACCATCGCT 240
QY 79 SerProLysGlyProValIleGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp 98
DB 241 TCCCGAAGGTCCCGTTTACCAGATGTACACCAACGTTGACAAAGACCTGGTTGGTGG 300
QY 99 ProAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 118
DB 301 CAGGCTCCGACGGTTCCTGTCACCGCGTTCGACCTGCGGTCTCTCCGACCTGTAC 360
QY 119 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 138
DB 361 CTGTTACCGCTCAGCTGACGTTATCCCGGTTCGTCGTCGTGACTCCGCTGGTTC 420
QY 139 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLysCys 158
DB 421 CTGCTGTCCCGCGTCCGATCTCTACCTGAAGGTTCCTCCGGTGGTCCGCTGTGTC 480
```

QY 159 ProAlaGlyHisAlaValGlyIlePheArgAlaValCysThrArgGlyValAlaLys 178
 Db 481 CCGGCTGGTCACGCTGTTGGTATCTTCGTCCTGCTTTCCACCGCTGGTGTCTAA 540
 QY 179 AlaValaspPheIleProValGluSerLeuGluThrThrMetArgSerPro 195
 Db 541 GCTGTTGACTTCATCCCGGTTGAATCCCTGGAAACCAACCATGCGTCCCG 591

RESULT 8

AAA73334
 ID AAA73334 standard; DNA; 594 BP.
 XX
 AC AAA73334;
 XX
 DT 19-DEC-2000 (first entry)
 XX
 DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #7.
 XX
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
 KW liver failure; liver cancer; mutant; mutein; ds.
 XX
 OS Hepatitis C virus.
 OS Synthetic.

XX Key Location/Qualifiers

FT CDS 1..594

FT /*tag= a

FT /product= "NS4A-NS3 fusion protein #7"

XX WO200040707-A1.

XX 13-JUL-2000.

XX 06-JAN-2000; 2000WO-US00345.

XX 08-JAN-1999; 99US-0115271.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

XX Wittekand M, Weinheimer S, Zhang Y, Goldfarb V;

XX WPI; 2000-465976/40.

DR P-PSDB; AAB15225.

XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 PT

XX Claim 26; Fig 17; 66pp; English.

XX The present sequence is the coding sequence for a mutated version of a
 CC fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
 CC protease enzymes. These proteins are both essential for the replication
 CC of the virus, acting to cleave its replicative proteins from the
 CC polypeptide produced from the HCV genome. Inhibitors of the two proteins
 CC should be effective as antiviral treatments of HCV infection. This is
 CC useful as HCV can lead to chronic liver disease such as cirrhosis, liver
 CC failure and liver cancer. The present invention concerns a number of NS3
 CC mutants and NS3-NS4A fusion proteins which can be used to identify
 CC inhibitors of this type, as well as enabling structural studies of the
 CC protease and protease-inhibitor complexes. The protein produced from this
 CC sequence contains the alpha-helix0-7 variant.

XX Sequence 594 BP; 105 A; 192 C; 151 G; 146 T; 0 other;

Alignment Scores:

Pred. No.: 1.08e-75 Length: 594
 Score: 929.00 Matches: 183
 Percent Similarity: 94.42% Conserved: 3
 Best Local Similarity: 92.89% Mismatches: 9
 Query Match: 90.99% Indels: 2
 DB: 21 Gaps: 1

US-09-965-594-12 (1-195) x AAA73334 (1-594)
 QY 1 MetLysLysLysGlySerValValIleValGlyArgIleValLeuasnGly-----Ala 18
 Db 1 ATGAAAAAAGATCCGTTGTTATCGTCGGCGGTATCAACCTGTCGGTGACACCGCT 60
 QY 19 TTTAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlnThrSerGlnThrGlyArgGasp 38
 Db 61 TACGCTCAGCAGACTCGAGGTGAGCAGGTACCCAGAAAGACCTCCACACCGGTGAC 120
 QY 39 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAla 58
 Db 121 AAAAAACAGGTGAAGGTGAAGTTGATGCTTCCACCGCTACCCAGACCTTCTCGGCT 180
 QY 59 ThrCysIleAsnGlyValCysTrrThrValTyHisGlyAlaGlyThrArgThrIleAla 78
 Db 181 ACCTCCATCAACGGTTCGTGTGACCGGTTTACCACCGGTGCTGTTACCCGCTCGCT 240
 QY 79 SerProLysGlyProValIleGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr 98
 Db 241 TCCCGAAAGTCCGGTTACCCAGATGTACCAACGTTGACAAAGACCTGGTGGTGG 300
 QY 99 ProAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 118
 Db 301 CAGGCTCCGACAGGTTCCTGTCACCGGTGACCTGCGGTTCCTCCGACCTGTAC 360
 QY 119 LeuValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySer 138
 Db 361 CTGTTTACCGGTCCAGCTGAGCTTATCCCGGTCGTGCTGCTGCTGCTGCTGCTGCT 420
 QY 139 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 158
 Db 421 CTGCTGTCCTCCGCGCTCCGATCTCTACTGAAAGGTTCTCCGCTGCTGCTGCTGCT 480
 QY 159 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys 178
 Db 481 CCGGCTGTCACGCTGTTGGTATCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTAAA 540
 QY 179 AlaValaspPheIleProValGluSerLeuGluThrThrMetArgSerPro 195
 Db 541 GCTGTTGACTTCATCCCGGTTGAATCCCTGGAAACCAACCATGCGTCCCGCG 591

RESULT 9

ABA95615
 ID ABA95615 standard; DNA; 12734 BP.
 XX
 AC ABA95615;
 XX
 DT 21-MAR-2002 (first entry)
 XX
 DE Chimeric BVDV/HCV NS3-wt sequence.
 XX
 KW Pestivirus; Npro; protease; NS3; screening; ds.
 XX
 OS Chimeric - Bovine viral diarrhea virus.
 OS Chimeric - Hepatitis C virus.
 XX
 PN US6326137-B1.
 XX
 PD 04-DEC-2001.
 XX
 PF 25-JUN-1999; 99US-0344456.
 XX
 PR 25-JUN-1999; 99US-0344456.
 XX
 PA (SCHE) SCHERING CORP.
 XX
 PI Hong Z, Lai VCH, Lau JYN;
 XX
 DR WPI; 2002-121103/16.
 XX
 PT Nucleic acid construct encoding chimeric Hepatitis C Virus (HCV)

PT pestivirus genome where the Npro protease gene is replaced with NS3
PT protease gene, useful for in vivo screening of compounds which inhibit
XX HCV infection -
XX
XX Example 2; Columns 17-28; 20pp; English.
XX
CC The present invention relates to a nucleic acid construct encoding a
CC chimeric Hepatitis C virus (HCV)-pestivirus genome. The construct
CC comprises a pestivirus genome where a Npro pestivirus protease gene is
CC replaced with a gene encoding a functional HCV NS3 protease. Furthermore,
CC each junction site recognised by the Npro protease is replaced with a
CC junction site recognised by the HCV NS3 protease. The construct is useful
CC for screening compounds that inhibit HCV in vivo by inhibiting HCV
CC protease, where screening may be in cell culture or in an animal model.
CC The present sequence is a chimeric clone of BVDV (bovine viral diarrhoea
CC virus)/HCV NS3-wt, which was used to illustrate the present invention.
XX
SQ Sequence 12734 BP; 4032 A; 2604 C; 3295 G; 2803 T; 0 other;

Alignment Scores:
Pred. No.: 5.05e-74 Length: 12734
Score: 928.50 Matches: 180
Percent Similarity: 94.87% Conservativity: 5
Best Local Similarity: 92.31% Mismatches: 5
Query Match: 90.94% Indels: 5
DB: 24 Gaps: 1

US-09-965-594-12 (1-195) x ABA95615 (1-12734)
QY 5 GlySerValIleValGlyArgIleValLeuAsnGly-----AlaTyr 19
DB 413 GGTAGTGTCTTATGTTGGTAGATGTTTATCTGGTAGTGGTAGTATCACGGCGTAC 472
QY 20 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 39
DB 473 GCCAGCAGAGAGAGCCCTCTAGGGTGTAGATCACAGTCTGACTGGCCGGACAAA 532
QY 40 AsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAlaThr 59
DB 533 AACCAAGTGGAGGTGAGGTCCAGATGCTGCTCACTGCTACCCAAACCTTCTCGCAACG 592
QY 60 CysIleAsnGlyValCysTyrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 79
DB 593 TGCATCAATGGGTATGCTGGAGTGTCTACACGGGCGGACGAGGACCATCGCATCA 652
QY 80 ProLysGlyProValIleGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrPro 99
DB 653 CCCAAGGTCTCTCATCCAGATGTATACCAATGTGGACCAACCTTGTGGCTGGCCC 712
QY 100 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 119
DB 713 GCTCTCAAGGTTCCCGCTCATTGACACCTCGACCTGCGGCTCCTCGGACCTTTACCTG 772
QY 120 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 139
DB 773 GTTACAGGACGCGGACGTCATTCCTCGCGCGGAGGTGATACAGGGGTAGCGCTG 832
QY 140 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 159
DB 833 CTTTCGCGCGCGCCATTCTCTACCTAAAGGCTCTCTCGGGGGTCTGCTGTGTGCCCC 892
QY 160 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 179
DB 893 GCGGACACGCGGTGGGCTTATTGAGGCGCGGTGTGCACCGGTGGAGTGGCAAGGCG 952
QY 180 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 194
DB 953 GTGGACTTTATCCCTGTGGAGAACCTTAGACACACCATGATGATCC 997

RESULT 10
ABX15706
ID ABX15706 standard; DNA: 612 BP.
XX

AC ABX15706;
XX
XX 28-MAR-2003 (first entry)
XX
XX Anti-viral synthetic prototoxophore associated DNA sequence.
XX
XX Hepatitis C; ds; viral prototoxophore; anti-viral; tumour;
KW virus; infection; antitumour; toxophore; human immunodeficiency virus;
KW HIV infection; herpes simplex virus; HSV; rhinovirus; NS3 protease.
XX
XX Unidentified.
XX
XX WO200287500-A2.
XX
XX 07-NOV-2002.
XX
XX 26-APR-2002; 2002WO-US13223.
XX
XX 27-APR-2001; 2001US-286893P.
XX
XX (NEWB-) NEWBIOTICS INC.
XX
XX Cathers BE, Neuteboom STC, Shepard HM;
XX
XX WPI; 2003-167102/16.
XX
XX Novel synthetic viral prototoxophore for treating viral infections, has
PT toxin moiety incorporated into substrate domain specific for viral
PT enzyme, bound and modified by viral enzyme to get converted into
PT toxophore -
XX
XX Example 1; Page 62; 66pp; English.
XX
XX This invention relates to a novel synthetic viral prototoxophore
CC comprising a toxin moiety operatively incorporated into a substrate
CC domain specific for a viral enzyme. This prototoxophore may be bound
CC and modified by the viral enzyme thus converting it to a toxophore.
CC Also disclosed in the invention is a method for enhancing the anti-viral
CC effect of an antiviral agent, this method comprises contacting a cell,
CC infected with a virus or is susceptible to infection, with a
CC prototoxophore. The invention further comprises an assay to identify
CC anti-viral agents, comprising contacting an infected cell with a
CC candidate agent and comparing the ability of the agent to inhibit the
CC growth or infectivity of the virus in the cell. The prototoxophores
CC of the invention may have virucide or antitumour activity. The
CC inhibiting viral infectivity, by contacting a cell (e.g. lymphocyte,
CC nerve cell, connective tissue cell, muscle cell or hepatocyte) which is
CC infected with a virus or is susceptible to infection with a virus, with
CC an effective amount of the prototoxophore. The cells are cell lines
CC adapted to long term continuous culture or isolated from a subject.
CC The prototoxophore is also useful for ameliorating the severity of a
CC viral infection in a subject, where the virus is selected from human
CC immunodeficiency virus (HIV), herpes simplex virus (HSV), rhinovirus and
CC hepatitis virus, by administering an effective amount of the
CC prototoxophore to the subject. The prototoxophores of the invention are
CC also useful for treating tumours. The present sequence represents an
CC antiviral prototoxophore associated DNA sequence, this sequence is
CC described as a recombinant NS3/NS4 fusion protein in example 1 of
CC the invention although it is clearly not a protein sequence.
XX
SQ Sequence 612 BP; 120 A; 171 C; 191 G; 130 T; 0 other;

Alignment Scores:
Pred. No.: 1.01e-74 Length: 612
Score: 918.50 Matches: 179
Percent Similarity: 94.36% Conservativity: 5
Best Local Similarity: 91.79% Mismatches: 6
Query Match: 89.96% Indels: 5
DB: 25 Gaps: 1

US-09-965-594-12 (1-195) x ABX15706 (1-612)

QY 72 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 91
 Db 1107 GCGGAAGAGGACCATCGGTACCCAGGGTCTGTATCCAGATGTATACCAATGTA 1166
 QY 92 AspLysAspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThr 111
 Db 1167 GACCAAGACCTTGTGGCTGGCCGCTCCCAAGGTAGCCGCTCATGTACACCTGCAC 1226
 QY 112 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 131
 Db 1227 TGGGCTCTCGGACCTTACCTGGTACAGGACGACGCGATGTCATTCGTCGCGG 1286
 QY 132 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 151
 Db 1287 CGGGGTGATAGCAGGGGACCCCTGCTGTCCGCCCGGCCCATTTCTTACTTGAAGGCTCC 1346
 QY 152 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 171
 Db 1347 TCGGGGGTCCGCTGTGTGCCCGGGGCGACGCCGTGGGCATATTTAGGGCCGCGTG 1406
 QY 172 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 191
 Db 1407 TGCACCGTGGAGTGGCTAAGGCGGTGGACTTTATCCTGTGGAGAACCCTAGAGACAACC 1466
 QY 192 MetArgSerPro 195
 Db 1467 ATGAGTCCCG 1478
 RESULT 12
 ID AAN90327 standard; cDNA: 5360 BP.
 AC AAN90327;
 XX
 XX 25-MAR-2003 (updated)
 DT 11-NOV-1989 (first entry)
 XX
 XX Hepatitis C virus composite probe.
 XX Hepatitis C virus; composite cDNA; probe: vaccine.
 XX Pan troglodytes.
 Key Location/Qualifiers
 CDS 3..5360
 /*tag= a
 GB2212511-A.
 26-JUL-1989.
 18-NOV-1988; 88GB-0027024.
 18-NOV-1987; 87US-0122714.
 30-DEC-1987; 87US-0139886.
 26-FEB-1988; 88US-0161072.
 26-OCT-1988; 88US-0263584.
 (CHIR) CHIRON CORPORATION.
 Houghton M, Choo QL, Kuo G;
 WPI: 1989-215054/30.

Hepatitis C virus gene - used for prodn. of polynucleotide probes,
 polypeptide(s) and antibodies for diagnosis, prevention and treatment
 of infection.

Disclosure; Fig. 26; 174pp; English.

The sequence shows the composite cDNA sequence derived from the aligned
 hepatitis C virus (HCV) cDNA's in clones 141, 11b, 7f, 7e, 8h, 33c, 40b,
 37b, 35, 36, 81, 32, 33b, 25c, 14c, 8f, 33f, 33g and 39c. The cDNA

CC encodes antigens which react with antibodies in patients with non-A
 CC non-B hepatitis (NANBH). The cDNA can be used to design probes, or to
 CC synthesize polypeptides, which can be used to diagnose HCV-induced NANBH,
 CC to raise antibodies for immunoassay or treatment, or to produce
 CC vaccines. See also AAP90158, AAN90303-26, and AAN90328-36.
 CC (updated on 25-MAR-2003 to correct PR field.)
 XX

Sequence 5360 BP; 1060 A; 1622 C; 1532 G; 1145 T; 1 other;

Alignment Scores:
 Pred. No.: 2,18e-71 Length: 5360
 Score: 894.50 Matches: 175
 Percent Similarity: 89.71% Conservative: 8
 Best Local Similarity: 85.78% Mismatches: 10
 Query Match: 87.61% Indels: 11
 DB: 10 Gaps: 2

US-09-965-594-12 (1-195) x AAN90327 (1-5360)

QY 3 LysLysGlySerValValIleValGly-----ArgIleValLeuAsnGly----- 17
 Db 867 CGCAGGGCGGGGAGATACCTGCTCGGGCCAGCGCATGGAATGCTCCCAAGGGGTGGAGG 926
 QY 18 -----AlaTyrAlaGlnGlnThrArgGlyGluGluCysGlnGlu 31
 Db 927 TTGCTGGCCCATCACGGCGTACGCCAGCAGACAAAGGGGCTCCTAGGGTGCATAATC 986
 QY 32 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 51
 Db 987 ACCAGCTAACTGGCCGGGACAAAACCAAGTGGAGGTGAGTCCAGATTGTGTCACT 1046
 QY 52 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGly 71
 Db 1047 GCTGCCAAACCTTCTGCAACGTGCATCAATGGGTGTGCTGGACTGTCTACCAACGG 1106
 QY 72 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 91
 Db 1107 GCCGGAACGAGGACCATCGCGTACCCAAAGGTCTCTCATCCAGATGTATACCAATGTA 1166
 QY 92 AspLysAspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThr 111
 Db 1167 GACCAAGACCTTGTGGCTGGCCGCTCCGCAAGGTAGCCGCTCATTCACACCTGCAC 1226
 QY 112 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 131
 Db 1227 TCGGCTCTCGGACCTTTACCTGTGTACAGGACGCGCATGTCTCCCTGCGCGG 1286
 QY 132 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 151
 Db 1287 CGGGGTGATAGCAGGGGACCCCTGCTGTCCGCCCGGCCCATTTCTTACTTGAAGGCTCC 1346
 QY 152 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 171
 Db 1347 TCGGGGGTCCGCTGTGTGCCCGGGGACGCGGTGGGCATATTTAGGGCCGCGTG 1406
 QY 172 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 191
 Db 1407 TGCACCGTGGAGTGGCTAAGGCGGTGGACTTTATCCTGTGGAGAACCCTAGAGACAACC 1466
 QY 192 MetArgSerPro 195
 Db 1467 ATGAGTCCCG 1478

RESULT 13
 ID AAN92103 standard; DNA; 6905 BP.
 XX AAN92103;
 XX
 XX 25-MAR-2003 (updated)
 DT 02-MAR-1990 (first entry)
 XX
 XX Combined open reading frames of the hepatitis C virus (HCV) cDNAs from

CC It is a double-stranded nucleotide sequence of the open reading frame
CC (ORF) (tag a) extending through clones K9-1 to 15e of hepatitis C virus
CC (HCV) cDNA. It can be used to make oligomeric DNA hybridisation probes to
CC detect the presence of HCV nucleic acids in samples. The polypeptide(s)
CC it encodes could be used as immunoassay reagents and vaccines and to
CC generate antibodies useful in diagnosis and passive immunotherapy for
CC HCV infection/non-A, non-B hepatitis.
CC (Updated on 25-MAR-2003 to correct PR field.)
XX (Updated on 25-MAR-2003 to correct PI field.)
SQ Sequence 7310 BP; 1491 A; 2217 C; 2058 G; 1540 T; 4 other;

Alignment Scores:
Pred. No.: 3.19e-71 Length: 7310
Score: 894.50 Matches: 175
Percent Similarity: 89.71% Conservative: 8
Best Local Similarity: 85.78% Mismatches: 10
Query Match: 87.61% Indels: 11
DB: 10 Gaps: 2

US-09-965-594-12 (1-195) x AAN92106 (1-7310)

QY 3 LysLysGlySerValValIleValGly-----ArgIleValLeuAsnGly----- 17
DB 1665 CCCAGGGCCGGGAGATACCTGCTCGGCCACGCCATGGATGGTCTCCAAAGGGGTGGAGG 1724
QY 18 -----AlaTyraAlaGlnGlnThrArgGlyGluGlyCysGlnGlu 31
DB 1725 TTGCTGGCGCCATCAGCGGTACGGCCAGACAAAGGGGCTCTAGGGTGCATAATC 1784
QY 32 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGlnIleValSerThr 51
DB 1785 ACCAGCTTAACCTGCGGGGCAAAACCAAGTGAGGGTCCAGTGTGTCAACT 1844
QY 52 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyHisGly 71
DB 1845 GCTGCCCAACCTCTCTGGCAACGTGCATCAATGGGTGTCTGGACTGTCTACCGGG 1904
QY 72 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyThrAsnVal 91
DB 1905 GCCGGAACGAGGACATCGGCTCACCCAAAGGGTCTGTCTCATCAGATGTATACCAATGA 1964
QY 92 AspLysAspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThr 111
DB 1965 GACCAAGACCTTGTGGCTGTGGCTGCGCGCTCCGCAAGGTAGCGGCTCATTTGACACCTGCAC 2024
QY 112 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 131
DB 2025 TCGCGCTCTCGGACCTTTACCTGTGCAGGAGCACGCCGATGTCTCCGTGGCGCGG 2084
QY 132 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyLeuLysGlySer 151
DB 2085 CGGGGTGATAGCAGGGGCAGCTGCTGTGCGCGCGGCCCATTTCTCTACTTGAAGGCTCC 2144
QY 152 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 171
DB 2145 TCGGGGGTTCGCTGTGTGCGCCCGGGGACGCCCTGGGCATATTTAGGGCGCGGTG 2204
QY 172 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 191
DB 2205 TGCACCCGTGAGTGGCTAAGCGGTGCAGTTTATCTCTGTGGAGAACCTAGACACAAC 2264
QY 192 MetArgSerPro 195
DB 2265 ATGAGGTCCCG 2276

RESULT 15

AAN90336

ID AAN90336 standard: DNA; 7310 BP.

XX AAN90336;

AC AAN90336;

XX AAN90336;

DT 25-MAR-2003 (updated)

DT 19-JUL-2001 (updated)
DT 01-NOV-1989 (first entry)
XX Composite hepatitis C virus (HCV) cDNA.
DE Hepatitis C virus; cDNA; clone 15e; clone K9-1; probe; vaccine; ds.
KW Pan troglodytes.
XX GB2212511-A.
XX 26-JUL-1989.
XX 18-NOV-1988; 88GB-0027024.
XX 18-NOV-1987; 87US-0122714.
PR 30-DEC-1987; 87US-0139886.
PR 26-FEB-1988; 88US-0161072.
PR 26-OCT-1988; 88US-0263584.
XX (CHIR) CHIRON CORPORATION.
XX Houghton M, Choo QL, Kuo G;
XX WPI; 1989-215054/30.
DR P-PSDB; AAP90288.
XX Hepatitis C virus gene - used for prodn. of polynucleotide probes,
PT polypeptide(s) and antibodies for diagnosis, prevention and treatment
PT of infection.
XX Disclosure; fig 47; 235pp; English.
XX The sequence shows a composite hepatitis C virus (HCV) cDNA, derived by
CC aligning clones K9-1 through 15e in 5'-3' direction. The cDNA
CC encodes antigens which react with antibodies in patients with non-A
CC non-B hepatitis (NANBH). The cDNA can be used to design probes, or to
CC synthesise polypeptides, which are used to diagnose HCV-induced NANBH,
CC to raise antibodies for immunoassay or treatment, or to produce
CC vaccines. See also AAP90288, and AAN90303-35.
CC (N.B. This record was resubmitted to correct errors in the sequence.)
CC (Updated on 25-MAR-2003 to correct PR field.)
XX Sequence 7310 BP; 1495 A; 2218 C; 2058 G; 1539 T; 0 other;
SQ Alignment Scores:
Pred. No.: 3.19e-71 Length: 7310
Score: 894.50 Matches: 175
Percent Similarity: 89.71% Conservative: 8
Best Local Similarity: 85.78% Mismatches: 10
Query Match: 87.61% Indels: 11
DB: 10 Gaps: 2
US-09-965-594-12 (1-195) x AAN90336 (1-7310)
QY 3 LysLysGlySerValValIleValGly-----ArgIleValLeuAsnGly----- 17
DB 1665 CCCAGGGCCGGGAGATACCTGCTCGGCCACGCCATGGATGGTCTCCAAAGGGGTGGAGG 1724
QY 18 -----AlaTyraAlaGlnGlnThrArgGlyGluGlyCysGlnGlu 31
DB 1725 TTGCTGGCGCCATCAGCGGTACGGCCAGACAAAGGGGCTCTAGGGTGCATAATC 1784
QY 32 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGlnIleValSerThr 51
DB 1785 ACCAGCTTAACCTGCGGGGCAAAACCAAGTGAGGGTCCAGTGTGTCAACT 1844
QY 52 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyHisGly 71
DB 1845 GCTGCCCAACCTCTCTGGCAACGTGCATCAATGGGTGTCTGGACTGTCTACCGGG 1904
QY 72 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyThrAsnVal 91
DB 1905 GCCGGAACGAGGACATCGGCTCACCCAAAGGGTCTGTCTCATCAGATGTATACCAATGA 1964
QY 92 AspLysAspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThr 111
DB 1965 GACCAAGACCTTGTGGCTGTGGCTGCGCGCTCCGCAAGGTAGCGGCTCATTTGACACCTGCAC 2024
QY 112 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 131
DB 2025 TCGCGCTCTCGGACCTTTACCTGTGCAGGAGCACGCCGATGTCTCCGTGGCGCGG 2084
QY 132 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyLeuLysGlySer 151
DB 2085 CGGGGTGATAGCAGGGGCAGCTGCTGTGCGCGCGGCCCATTTCTCTACTTGAAGGCTCC 2144
QY 152 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 171
DB 2145 TCGGGGGTTCGCTGTGTGCGCCCGGGGACGCCCTGGGCATATTTAGGGCGCGGTG 2204
QY 172 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 191
DB 2205 TGCACCCGTGAGTGGCTAAGCGGTGCAGTTTATCTCTGTGGAGAACCTAGACACAAC 2264
QY 192 MetArgSerPro 195
DB 2265 ATGAGGTCCCG 2276

```
Db 1905 GCCGGAACGAGGACCATCGCTACCCAAAGGTCTCTCATCCAGATGTATACCAATGTA 1964
Qy 92 AspLysAspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThr 111
Db 1965 GACCAAGACCTTGTGGGCTGGCCGCTCCGCAAGGTAGCCGCTCATTCACACCCCTGCAC 2024
Qy 112 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 131
Db 2025 TCGGGCTCTCGGACCTTACTGGTCACGAGGCACGCCGATGTCATTCCCGTGGCCGG 2084
Qy 132 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 151
Db 2085 CGGGGTGATAGCAGGGGCGAGCTGTGTGCGCCCGGCCCATTTCTACTTGAAGGCTCC 2144
Qy 152 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 171
Db 2145 TCGGGGGGTCCGCTGTGTGCCCCGCGGGGCACGCCGTGGGCATATTTAGGGCCCGGTG 2204
Qy 172 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 191
Db 2205 TGCACCCGTGGAGTGGCTAAGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACACC 2264
Qy 192 MetArgSerPro 195
Db 2265 ATGAGGTCCCCG 2276
```

Search completed: August 30, 2003, 19:47:47
Job time : 190.082 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 30, 2003, 19:20:43 : Search time 1890.92 Seconds
(without alignments)
2506.388 Million cell updates/sec

Title: US-09-965-594-12

Perfect score: 1021

Sequence: 1 MKKGSWVIGRVLNGAYA.....VAKAVDFIPVESLETTMRSP 195

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q/cgn2_1/USPTO.spool/US09965594/runat_29082003_151919_28322/app_query.fasta_1.2872
-DB=EST -OPT=fastap -SUFFIX=est -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09965594.ecgn_1_12630_@runat_29082003_151919_28322 -NCPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST: *
1: em_estba: *
2: em_esthum: *
3: em_estin: *
4: em_estnu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_hic: *
9: gb_est1: *
10: gb_est2: *
11: gb_hic: *
12: gb_est3: *
13: gb_est4: *
14: gb_est5: *
15: em_estfun: *
16: em_estom: *
17: em_gss_hum: *
18: em_gss_inv: *
19: em_gss_pln: *
20: em_gss_vrt: *
21: em_gss_fun: *
22: em_gss_mam: *
23: em_gss_mus: *
24: em_gss_pro: *
25: em_gss_rod: *
26: em_gss_phg: *
27: em_gss_vrl: *
28: gb_gss1: *

29: gb_gss2: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	106	10.4	984	10	BF304699	BF304699 601888252
C 2	103.5	10.1	1199	13	BQ892487	BQ892487 AGENCOURT
C 3	103	10.1	814	13	BU148820	BU148820 AGENCOURT
C 4	101.5	9.9	1403	13	BQ926101	BQ926101 AGENCOURT
C 5	100.5	9.8	844	12	BI198486	BI198486 602760491
C 6	100.5	9.8	1204	13	BQ881847	BQ881847 AGENCOURT
C 7	99	9.7	615	12	BQ001625	BQ001625 BJ001625
C 8	99	9.7	643	12	BJ024121	BJ024121 BJ024121
C 9	99	9.7	754	12	BJ016176	BJ016176 BJ016176
C 10	98.5	9.6	961	10	BF203316	BF203316 601865914
C 11	98	9.6	1141	11	AK080545	AK080545 MUS muscu
C 12	97.5	9.5	779	10	BF631437	BF631437 HVSMED001
C 13	97	9.5	1146	12	BM915803	BM915803 AGENCOURT
C 14	96	9.4	701	10	BF863244	BF863244 963042CQ2
C 15	96	9.4	846	10	BF182274	BF182274 601804028
C 16	95.5	9.4	901	10	BF307233	BF307233 601891502
C 17	95.5	9.4	930	13	BU169585	BU169585 AGENCOURT
C 18	95	9.3	407	9	AW785806	AW785806 117260_MA
C 19	95	9.3	958	10	BG420860	BG420860 602452062
C 20	95	9.3	1384	29	CC221189	CC221189 CH261-183
C 21	95	9.3	1433	12	BM803824	BM803824 AGENCOURT
C 22	94.5	9.3	641	9	AU127824	AU127824 AU127824
C 23	94.5	9.3	938	13	BQ894657	BQ894657 AGENCOURT
C 24	94	9.2	649	10	BE289911	BE289911 601089126
C 25	94	9.2	940	14	CB993468	CB993468 AGENCOURT
C 26	94	9.2	1283	13	BQ709745	BQ709745 AGENCOURT
C 27	93.5	9.2	701	14	CD262790	CD262790 PSMA019xE
C 28	93.5	9.2	832	10	BG387051	BG387051 602454749
C 29	93.5	9.2	905	13	BQ542842	BQ542842 AGENCOURT
C 30	93.5	9.2	993	9	AL555424	AL555424 AGENCOURT
C 31	93.5	9.2	1001	13	BQ928211	BQ928211 AGENCOURT
C 32	93	9.1	964	12	BI196460	BI196460 602755151
C 33	92.5	9.1	556	14	CB216999	CB216999 NISC_nq11
C 34	92.5	9.1	715	9	AU125614	AU125614 AGENCOURT
C 35	92.5	9.1	846	13	BQ540812	BQ540812 AGENCOURT
C 36	92.5	9.1	866	13	BX451426	BX451426 BX451426
C 37	92.5	9.1	881	14	CD105862	CD105862 AGENCOURT
C 38	92.5	9.1	929	13	BQ672290	BQ672290 AGENCOURT
C 39	92.5	9.1	947	13	BQ536872	BQ536872 AGENCOURT
C 40	92.5	9.1	979	13	BQ673186	BQ673186 AGENCOURT
C 41	92.5	9.1	1008	12	B1755608	B1755608 603027112
C 42	92.5	9.1	1169	12	BM548430	BM548430 AGENCOURT
C 43	92.5	9.1	1291	10	BE622016	BE622016 601440668
C 44	92.5	9.1	1384	13	BQ919246	BQ919246 AGENCOURT
C 45	92	9.0	582	14	CB286751	CB286751 CMD45_C08

ALIGNMENTS

RESULT 1
BF304699/c
LOCUS 601888252F1 NIH_MGC_17 Homo sapiens CDNA clone IMAGE:4122276 5',
DEFINITION mRNA sequence.
ACCESSION BF304699
VERSION BF304699.1 GI:11251586
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 984)

	(..... :	Db	243 GAAGTGATAC	TGTTAGTGGAGGAT-----ACTGTCTGGAG 281
	63 GlyVal---CysIleThrValTyHisGlyAlaGlyThr-----ArgThrlleAla 78	QY		
	282 GGGTACTACTGCTGGAGGGGTACTATGCTGGAGGGTACTGTCCGGGAAGGGTACTGT 341	Db		
	79 SerProLysGlyProValIleGlMetTy-----ThrAsnValasPlys 93	QY		
	342 CGTGAGGGGGTACTGTCTCGAAAGGATACTGCTGGAANGTGATATCTATTAGTGGAG 401	Db		
	94 AspleuVal-----gltTrpProAlaprogInglySerArg 105	QY		
	402 GATACGTGCTGGAGTGGTACTGCTGCCGAGGATACTGTGCT-----GGGGGAGA 455	Db		
	106 SerLeuThrProCysThrCysGlySerSerAspleuTyrrLeuValThrArgHlsAlaasp 125	QY		
	456 TCTCTGGTCGAATGGCATGTGCCTGCCACTGCTGTTCAI-----CACCNNGAT 503	Db		
	126 ValIleProValArgArgGlyAspSerArgGlySerLeuLeuSerProArgProlie 145	QY		
	504 -----TCGTGGCTCGAGCTTGCCCTGAATTAATCTCTCAAACCCT 545	Db		
	146 SertyrLeuLysGlySerSerglyglyproLeu 156	QY		
	546 TTCCTITCCAAGGGTCTGACGCTTTCCCCTG 578	Db		
RESULT 4				
BQ326101/c				
LOCUS				
DEFINITION	BQ926101 1403 bp mRNA linear EST 20-AUG-2000			
DESCRIPTION	5' , mRNA sequence.			
ACCESSION	BQ926101			
VERSION	BQ926101.1 GI:22341132			
KEYWORDS	EST.			
SOURCE	Mus musculus (house mouse)			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;			
REFERENCE	I (bases 1 to 1403) NIH-MGC http://mgc.nci.nih.gov/ National Institutes of Health, Mammalian Gene Collection (MGC)			
AUTHORS	Contact: Robert Strausberg, Ph.D. Email: cgabbs@mail.nih.gov Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman, Ph.D.			
JOURNAL	CDNA Library Preparation: ResGen, Invitrogen Corp CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Cloned through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov plate: LLAM13798 row: j column: 07 High quality sequence stop: 101.			
COMMENT	Location/Qualifiers 1..1403 /organism="Mus musculus" /mol_type="mRNA" /db_xref="taxon:10090" /clone="IMAGE:6335718" /lab_host="DH10B (phage-resistant)" /clone_lib="NIH_MGC_130" /note="Organ: otcysts; Vector: pCMV-SPORT6.1.ccd; Site.1: ECORV; Site.2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.95 kb. Constructed by ResGen, Invitrogen Corp. Note: This is a NIH_MGC Library."			
BASE COUNT	297 a 521 c 237 g 345 t 3 others			
ALIGNMENT SCORES:	18.5 Length: 1403			
Pred. No.:				

Score: 101.50 Matches: 46
Percent Similarity: 36.75% Conservative: 15
Best Local Similarity: 27.71% Mismatches: 68
Query Match: 9.94% Indels: 37
DB: 13 Gaps: 7

US-09-965-594-12 (1-195) x BQ926101 (1-1403)

```
Qy 4 LysGlySerValValIleValGlyArgIleValLeuAsnGlyAlaTyrAlaGlnGlnThr 23
Db 1381 AGAGGTGTGTCANCGGTTCAGACAGGTCCGCGACACTCGAGCGTCCGCCAGAGACT 1322
Qy 24 ---ArgGlyGluGluGlyCysGlnGlnThrSerGlnThrGlyArgAspLysAsnGlnVal 42
Db 1321 TGTGGGGCGGTGGTGCATACCCCGGTTCGATCGAGGTTCAGCGCGCTTGTATACA 1262
Qy 43 GluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAlaThrCysIleAsn 62
Db 1261 GAGGGGAAA-----CAG 1250
Qy 63 GlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLys--G 82
Db 1249 GGGGTA---TGGTTATCACGGGTGGGGCAGGTACT-----TCCCTAAAGCG 1205
Qy 82 lyProValIleGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpProAlaProG 102
Db 1204 CGCGGTGGCGAGTATATACCGCGGAGTGCAGCGCAGCGCGGTGGAAGTTGACC 1145
Qy 102 InGlySerAlaGlySerLeuThrProCysThrCysGlySerSerAspLeuTyrLeuValThr 122
Db 1144 AA---GAGAGCGACCTGACGCCCTCCCTGTGGGNTGTCGATATACAAATGTGCAG 1088
Qy 122 rgHisAlaAspValIleProValArgArgGlyAsp----- 134
Db 1087 GGCACGGTGATGTGTTACTACGGCGGAGACCGGCTCCACGGCGCTCTCTAACAGACGC 1028
Qy 135 -----SerArgGlySer-----LeuLeuSerProArgProIleSerTyrLeuLysG 150
Db 1027 CGCGCTCCCGCGGCAACAGGTGTAATATCATATCATATCGCGGCGGGATTTGCGCATTCGCGCGG 968
Qy 150 lySerSerGlyGly 154
Db 967 GAGAGCGCGCGGT 954
```

RESULT 5
BI198486/c 844 bp mRNA linear EST 10-JUL-2001
LOCUS 602760491F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4895638 5',
DEFINITION mRNA sequence.

ACCESSION BI198486
VERSION BI198486.1 GI:14653507
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 844)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs.r@mail.nih.gov
Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LICM1781 row: f column: 23
High quality sequence stop: 794.
Location/Qualifiers

FEATURES
source
1..844 /organism="Homo sapiens"

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4895638"
/tissue_type="neuroblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_19"
/note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
ECORI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 203 a 230 c 224 g 187 t
ORIGIN

Alignment Scores:

Pred. No.:	12.1	Length:	844
Score:	100.50	Matches:	50
Percent Similarity:	35.56%	Conservative:	35
Best Local Similarity:	20.92%	Mismatches:	78
Query Match:	9.84%	Indels:	77
DB:	12	Gaps:	11

US-09-965-594-12 (1-195) x BI198486 (1-844)

```
Qy 12 ArgIleValLeuAsnGlyAlaTyrAlaGlnGlnThrArgGlyGluGlyCysGlnGlu 31
Db 791 CGGTGATGATAGTACGAGTTCAGTTCGACAGTTCAGTACCCATCTCGACAGAC 732
Qy 32 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 51
Db 731 ACCCTGCGCAACATTCGCGACACCCCTGCCAGGAGCGAGAAATGCGAGAGTTCTCTGTG 672
Qy 52 AlaAlaGlnThrPhe---LeuAlaThrCys-----IleAsnGlyValCysTrpThrVal 68
Db 671 ATATCAGGACATTTTACAGGTTTGTAGATGCTGCCATTTGTCGAACACCTCTGTGG--- 618
Qy 69 TyrHisGlyAlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyr 88
Db 617 -----ATGACCACCCCAAGGAGAGGAGGAGATGTTGAGCATGTTTC 576
Qy 89 ThrAsnValAspLysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThr 108
Db 575 AGCAGCGTG-----GCTTCGCTGGCT 555
Qy 109 ProCysThrCysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIlePro 128
Db 554 CCACCTTTGCTCCAGGTCTTGATCAGTGCACATCATCTCAGGATTTCAATGGTGCCTTC 495
Qy 129 ValArgArgGlyAspSer-----ArgGlySerLeuLeuSerProArgPro 144
Db 494 GGAGATTTAGTGGTGATACCTAAACCTGGAAAAGGAGGCTCTTCGCGCCGACACCA 435
Qy 145 Ile-----SerTyrLeuLysGlySerSerGlyGly----- 154
Db 434 GTGTTCTGGGTGGCACAGTGCATTCAGCATGGGCAATGGCAGCAGCAGCGGTTCAGCAG 375
Qy 155 -----ProLeuLeuCysPro----- 159
Db 374 CTGGCACCTTATTTGGCCAGCAACATGTCCTCATCTCATCTGAGGTCTCTCTTTGGTGAACA 315
Qy 160 -----AlaGlyHisAlaValGlyIlePhe---ArgAlaAlaVal----- 171
Db 314 CAAGAGCCACATTCCTCCCGGATATGAGGAGCAGGT-TTCTCCAGAGCTGGGTGTTTCC 256
Qy 172 ---CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSer----- 187
Db 255 AGTGCCCTCGGATGGCTTGGCATCATGGTGTCTTCTGCCCATCAGCAGCAGCCTTC 196
Qy 188 -----LeuGluThrThrMetArgSerPro 195
Db 188 -----LeuGluThrThrMetArgSerPro 195
```

```

Db      195 CCGGAGGACATGCGGATCTCTCTGATCTGCTTGGAGCCACATGTCTGCTCC 139
RESULT 6
BQ881847/c
LOCUS   BQ881847      1204 bp      mRNA      linear      EST 16-AUG-2002
DEFINITION AGENCOURT_8712410 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6295246
5', mRNA sequence.
ACCESSION BQ881847
VERSION   BQ881847.1 GI:22273855
KEYWORDS EST.
SOURCE    Homo sapiens (human)
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1204)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cyapbs-r@mail.nih.gov
          Tissue Procurement: DCTD/DTF
          cDNA Library Preparation: Rubin Laboratory
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LCM2501 row: c column: 23
          High quality sequence stop: 322.
          Location/Qualifiers
            location=1..1204
            organism="Homo sapiens"
            mol_type="mRNA"
            db_xref="taxon:9606"
            clone="IMAGE:6295246"
            tissue_type="melanotic melanoma, cell line"
            lab_host="DH10B (phage-resistant)"
            notes="Organ: skin; Vector: pOTB7; Site:1: XhoI; Site:2:
            EcoRI; cDNA made by oligo-dT priming. Directionally cloned
            into EcoRI/XhoI sites using the following 5' adaptor:
            GGCACAGAG(G). Library constructed by Ling Hong in the
            laboratory of Gerald M. Rubin (University of California,
            Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
            Superscript II RT (Life Technologies). Note: this is a
            NIH_MGC Library."
BASE COUNT 228 a 430 c 265 g 280 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 19 Length: 1204
Score: 100.50 Matches: 41
Percent Similarity: 36.02% Conservative: 17
Best Local Similarity: 25.47% Mismatches: 58
Query Match: 9.84% Indels: 45
DB: 13 Gaps: 6

US-09-965-594-12 (1-195) x BQ881847 (1-1204)

Qy      10 ValGlyArgIleValLeuAsnGlyAlaTyrAlaGlnGlnThrArgGly----- 25
      111 111 111111 111 111
Db      773 GTGCCAAGGGGTGTGTCACGGGGCAGTATATGCGCCAAACCGTGGTGGCAACGCC 714
Qy      26 -----GluGluGlyCysGlnGluThrSerGlnThrGlnThrArgAspLysAsn 40
      111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db      713 GGATAAAGGGCGCAACAGCGGGCGGCTGGCTAATTAAAGAGGGAGTAAACAAC 654
Qy      41 GlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAlaThrCys 60
      111111 111111 111111 111111 111111 111111 111111 111111
Db      653 CGCGGGGAGGGCGCAAGGAATATATCTGGAGGTGTCCGGGGGGCCATAGGGAGGGA 594
Qy      61 IleAsnGlyValCysTrp-----ThrValTyrHisGlyAlaGlyThrArgThrIle 77
      ::::: 111111 111111 111111 111111 111111 111111 111111

```

```

Db      593 CACGACGGAAGATGTTGGTTCAGGATGCGGGGAAAGAGCGGGGGGAGGCTATA 534
Qy      78 AlaSerProLysGlyProValIleGlnMetTyrThrAsnValAspLysAspLeuValGly 97
      111 111 111111 111111 111111 111111 111111 111111 111111 111111
Db      533 GGGGGGGAGGAGGTGGTAAAT-----ATAGTGGGT 501
Qy      98 -----TTPProAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySer 114
      111111 111111 111111 111111 111111 111111 111111 111111 111111
Db      500 GTATAACGGTGGCGGGGCTCGTGGTCTCG----- 468
Qy      115 SerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgGlyAsp 134
      111111 111111 111111 111111 111111 111111 111111 111111 111111
Db      467 -----GTCTCCGACCCCTTGGCGGTTATCCCGGGTCCGGAG 423
Qy      135 SerArgGlySerLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGly 154
      111111 111111 111111 111111 111111 111111 111111 111111 111111
Db      422 ACGGAGGGGGGATGTATCG-----AAGGAGGGGGGGGGA 384
Qy      155 Pro 155
      111
Db      383 CCG 381
      111
RESULT 7
BQ001625/c
LOCUS   BQ001625      615 bp      mRNA      linear      EST 05-DEC-2001
DEFINITION BQ001625 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA025C02 5',
mRNA sequence.
ACCESSION BQ001625
VERSION   BQ001625.1 GI:17364516
KEYWORDS EST.
SOURCE    Oryzias latipes (Japanese medaka)
ORGANISM Oryzias latipes
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
          Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
          Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE 1 (bases 1 to 615)
AUTHORS   Kohara,Y., Shin-i,T., Kimura,T., Marita,T., Jindo,T. and Takeda,H.
TITLE     Medaka EST Project in Takeda's lab
JOURNAL   Unpublished
COMMENT   Contact: Tadasu Shin-i
          Center for Genetic Resource Information
          National Institute of Genetics
          1111 Yata, Mishima, Shizuoka 411-8540, Japan
          Tel: 81-559-81-6856
          Fax: 81-559-81-6855
          Email: tshiniegenes.nig.ac.jp.
FEATURES
          source
            location/Qualifiers
              1..615
              /organism="Oryzias latipes"
              /mol_type="mRNA"
              /strain="Hd-rR"
              /db_xref="taxon:8090"
              /clone="MF01SSA025C02"
              /sex="mixture of female and male"
              /tissue_type="whole embryo"
              /dev_stage="segmentation stage 20 - 25"
              /clone_lib="MF01SSA cDNA"
BASE COUNT 140 a 166 c 165 g 144 t
ORIGIN
Alignment Scores:
Pred. No.: 11.3 Length: 615
Score: 99.00 Matches: 42
Percent Similarity: 33.77% Conservative: 9
Best Local Similarity: 27.81% Mismatches: 50
Query Match: 9.70% Indels: 50
DB: 12 Gaps: 7

US-09-965-594-12 (1-195) x BQ001625 (1-615)

Qy      39 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAla 58
      111111 111111 111111 111111 111111 111111 111111 111111

```



```

Db      511 AAAATGACGTAGAACCAACACACACAGATCCACACACATGTTCTGTTCTACGGGCT 452
QY      59 ThrCysIleAsnGlyValCysTrpThrValTyHisGlyAlaGlyThrArgThrIleAla 78
Db      451 -----TGTGGAGAACCTATCACAGTTCTGCTTTAGACCAACGGCA 410
QY      79 SerProLys-----GlyProValIleGlnMetTyThrAsnValAspLys 93
Db      409 GCTCTCGCGCGGAGAGCTCTCTGGCCAGTTGTG----- 374
QY      94 AspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrPro----- 109
Db      373 -----ACTGTGGAGAACCAAGAGCGTCACCCGGAGCTGTAGG 335
QY      110 -----CysThrCysGlySerAspLeuTyLeuValThrArg----- 122
Db      334 CTGACGGATGCGGATGTGGCTCTGCT-----TTGGTTCTCTGCTCTCGATCA 284
QY      123 -----HisAlaAspValIleProValArgArgArgGlyAspSer 135
Db      283 TCTTCTCCTACCTGACCTTCCACATCCAGGTGTCGCCAGCGCTGCTGACGGGTGATGG 224
QY      136 ArgGlySerLeuLeuSerProArg-----ProIleSerTyLeuLysGlySer 152
Db      223 ACAGCGCGGACAGGACAGTCGGGGTGAATCTCTGCAGGACGCTTCTACGGCGGATCA 164
QY      153 GlyGlyProLeuLeuCysProAlaGlyHisAla 163
Db      163 GGAGGACCGACTCGCTGCAGAGCGCTCTGCTGCA 131

RESULT 8
BJ024121
LOCUS      BJ024121 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA143D12 3',
DEFINITION mRNA sequence.
ACCESSION BJ024121
VERSION    BJ024121.1 GI:17377389
KEYWORDS   EST.
SOURCE     Oryzias latipes (Japanese medaka)
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
            Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE  1 (bases 1 to 643)
AUTHORS    Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
TITLE      Medaka EST Project in Takeda's lab
JOURNAL    Unpublished
COMMENT    Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
            Location/Qualifiers
            1..643
            /organism="Oryzias latipes"
            /mol_type="mRNA"
            /strain="Hd-rR"
            /db_xref="taxon:8090"
            /clone="MF01SSA143D12"
            /sex="mixture of female and male"
            /tissue_type="whole embryo"
            /dev_stage="segmentation stage 20 - 25"
            /clone_lib="MF01SSA cDNA"

BASE COUNT 171 a 148 c 148 g 176 t
ORIGIN

Alignment Scores:
Pred. No.: 12 Length: 643
Score: 99.00 Matches: 42
Percent Similarity: 33.77% Conservative: 9

```

```

Best Local Similarity: 27.81% Mismatches: 50
Query Match: 9.70% Indels: 50
DB: 12 Gaps: 7
US-09-965-594-12 (1-195) x BJ024121 (1-643)

QY      39 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAla 58
Db      242 AAAAATGACGTAGAACCAACACACACATGTTCTGTTCTACGGGCT 301
QY      59 ThrCysIleAsnGlyValCysTrpThrValTyHisGlyAlaGlyThrArgThrIleAla 78
Db      302 -----TGTGGAGAACCTATCACAGTTCTCTGCTTTAGACCAACGGCA 343
QY      79 SerProLys-----GlyProValIleGlnMetTyThrAsnValAspLys 93
Db      344 GCTCTCGCGCGGAGAGCTCTCTGGCCAGTTGTG----- 379
QY      94 AspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrPro----- 109
Db      380 -----ACTGTGGAGAACCAAGAGCGTCACCCGGAGCTGTAGG 418
QY      110 -----CysThrCysGlySerAspLeuTyLeuValThrArg----- 122
Db      419 CTGACGGATGCGGATGTGGCTCTGCT-----TTGGTTCTCTGCTCTCTCGATCA 469
QY      123 -----HisAlaAspValIleProValArgArgGlyAspSer 135
Db      470 TCTTCTCCTACCTGACCTTCCACATCCAGGTGTCGCCAGCGCTGCTGACGGGTGATGG 529
QY      136 ArgGlySerLeuLeuSerProArg-----ProIleSerTyLeuLysGlySer 152
Db      530 AGAGCGCGGACAGCAGCAGTCGGGGTGAATCTCTGCAGGACGCTTCTACGGCGGATCA 589
QY      153 GlyGlyProLeuLeuCysProAlaGlyHisAla 163
Db      590 GGAGGACCGACTCGCTGCAGAGCGCTCTGCTGCA 622

RESULT 9
BJ016176
LOCUS      BJ016176 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA025C02 3',
DEFINITION mRNA sequence.
ACCESSION BJ016176
VERSION    BJ016176.1 GI:17376695
KEYWORDS   EST.
SOURCE     Oryzias latipes (Japanese medaka)
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
            Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE  1 (bases 1 to 754)
AUTHORS    Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
TITLE      Medaka EST Project in Takeda's lab
JOURNAL    Unpublished
COMMENT    Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
            Location/Qualifiers
            1..754
            /organism="Oryzias latipes"
            /mol_type="mRNA"
            /strain="Hd-rR"
            /db_xref="taxon:8090"
            /clone="MF01SSA025C02"
            /sex="mixture of female and male"
            /tissue_type="whole embryo"
            /dev_stage="segmentation stage 20 - 25"

FEATURES
source

```

FEATURES

JOURNAL MEDLINE PUBMED REFERENCE	Genome Res. 10 (10), 1617-1630 (2000) 20499374 11042159
AUTHORS	3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsu, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Todawa, K., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, K., Kira, A., and Hayashizaki, Y.
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL MEDLINE PUBMED REFERENCE	Genome Res. 10 (11), 1757-1771 (2000) 20530913 11076861
AUTHORS	4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fieschmann, M., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nakido, I., Pesole, G., Quackenbush, J., Schriml, L.M., Staib, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyooka, K., Wang, K.H., Weitz, C., Whittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S., and Hayashizaki, Y.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL MEDLINE PUBMED REFERENCE	Nature 409 (6821), 685-690 (2001) 21095660 11217851
AUTHORS	5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL MEDLINE PUBMED REFERENCE	Nature 420, 563-573 (2002) 6 (bases 1 to 1141)
AUTHORS	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koyas, K., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohnato, N., Okazaki, Y., Saito, R., Saitoh, K., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takada, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.
TITLE	Direct Submission
JOURNAL MEDLINE PUBMED REFERENCE	Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsic.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details.


```

Qy 120 lIhArGHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeuLe 140
Db 609 -ACTACACACACT--GTAATCCAGGAGGAACAGGNTGGAGGAACAGAGGACTCC---CT 556
Qy 140 uSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeu--LeuCysPr 159
Db 555 GACCCCACTCCC---TCGGTCCTAGCGGGGACACTCTCTCGGCCCACTCTCTCTGTCC 499
Qy 159 oAlaGlyHis---AlaValGlyIlePheArg-----AlaAlaValCysThrAr 174
Db 498 TAGTGGGCACCTCTCTCCAGGCACGACAGACTGTACTCTCCCTTTGGCCCTCTGCACCTCT 439
Qy 174 gGlyValAlaLys 178
Db 438 TGGGATGACTGAG 426

Search completed: August 31, 2003, 04:27:24
Job time : 1896.92 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run On: August 30, 2003, 17:42:58 ; Search time 44.6227 Seconds
(without alignments)
700.745 Million cell updates/sec

Title: US-09-965-594-14

Perfect score: 1032

Sequence: 1 MKKGSVVIGRINLSGDTA.....VAKAVDPVPESTETMRSP 197

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

- 1: /SIDSL1/qcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDSL1/qcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDSL1/qcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDSL1/qcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDSL1/qcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDSL1/qcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDSL1/qcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDSL1/qcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDSL1/qcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDSL1/qcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDSL1/qcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDSL1/qcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDSL1/qcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDSL1/qcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDSL1/qcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDSL1/qcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDSL1/qcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDSL1/qcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDSL1/qcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDSL1/qcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDSL1/qcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDSL1/qcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDSL1/qcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
- 24: /SIDSL1/qcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1032	100.0	197	21	AA15221 Hepatitis C virus
2	1015	98.4	197	21	AA15222 Hepatitis C virus
3	998	96.7	195	21	AA15220 Hepatitis C virus
4	993	96.2	197	21	AA15226 Hepatitis C virus
5	985	95.4	197	21	AA15223 Hepatitis C virus
6	973	94.3	197	21	AA15224 Hepatitis C virus
7	963	93.3	197	21	AA15225 Hepatitis C virus
8	959	92.9	195	21	AA15212 Hepatitis C virus
9	902.5	87.5	3011	14	AA40120 HCV genomic amino

10	901.5	87.4	1766	10	AA192041	Sequence encoded 1
11	901.5	87.4	1786	10	AA190158	Protein sequence 0
12	901.5	87.4	2261	10	AA190164	Peptide encoded by
13	901.5	87.4	2301	10	AA192047	Sequence encoded 1
14	901.5	87.4	2436	10	AA192050	Sequence encoded 1
15	901.5	87.4	2436	10	AA190288	Protein encoded by
16	901.5	87.4	2772	21	AA18540	HCV-1 polyprotein.
17	901.5	87.4	2816	14	AA1834009	HCV-1 polyprotein.
18	901.5	87.4	2894	16	AA170230	Composite hepatitis
19	901.5	87.4	2955	20	AA14975	Amino acid sequenc
20	901.5	87.4	2955	21	AA18541	Polyprotein encode
21	901.5	87.4	3011	13	AA121519	Compiled HCV sequ
22	901.5	87.4	3011	14	AA131621	Hepatitis C virus
23	901.5	87.4	3011	17	AA190931	Hepatitis C virus
24	901.5	87.4	3011	18	AA134480	HCV polyprotein.
25	901.5	87.4	3011	19	AA140038	HCV polyprotein.
26	901.5	87.4	3011	23	AA122049	Hepatitis C virus
27	901.5	87.4	3011	23	AA184597	HCV polyprotein la
28	900.5	87.3	3011	15	AA186995	Hepatitis C virus
29	899	87.1	182	21	AA15211	Hepatitis C virus
30	899	87.1	609	15	AA151170	Hepatitis C virus
31	899	87.1	631	18	AA131884	A nonstructural pr
32	899	87.1	686	23	AA18689	HCV-1 NS3/4a mutan
33	899	87.1	686	23	AA176377	Hepatitis C virus
34	899	87.1	686	24	AB172261	Hepatitis C virus
35	898.5	87.1	2435	13	AA125135	HCV-1 NS3/4a conf
36	898.5	87.1	2436	13	AA128582	HCV amino acid seq
37	897.5	87.0	2772	11	AA108123	Hepatitis C virus
38	897	86.9	631	20	AA193482	HCV NS3 protein.
39	896.5	86.9	3011	19	AA177397	Hepatitis C virus
40	896.5	86.9	3011	24	AB171460	Amino acid sequenc
41	896.5	86.9	3012	23	AA109289	Hepatitis C virus
42	895	86.7	632	23	AA121847	Hepatitis C virus
43	895	86.7	632	23	AA119005	Hepatitis C virus
44	895	86.7	686	23	AA121837	Hepatitis C virus
45	895	86.7	686	23	AA121838	Hepatitis C virus

ALIGNMENTS

RESULT 1
AA15221
ID AA15221 standard; protein; 197 AA.
XX
AA15221;
XX
19-DEC-2000 (first entry)
XX
Hepatitis C virus NS4A-NS3 fusion protease #3.
DE
Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW
liver failure; liver cancer; mutant; mutain.
XX
Hepatitis C virus.
OS
Synthetic.
XX
WO200040707-Al.
XX
13-JUL-2000.
XX
06-JAN-2000; 2000WO-0500345.
XX
08-JAN-1999; 99US-0115271.
XX
(BRIM) BRISTOL-MYERS SQUIBB CO.
XX
Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX
WPI: 2000-465976/40.
DR
N-PSDB; AAA73330.
XX
Modified hepatitis C virus (HCV) NS3 protease comprising at least 1

PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 PT -
 XX
 PS Claim 23; Fig 13; 66pp; English.
 XX
 CC The present sequence is a mutated version of a fusion protein created
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
 CC proteins are both essential for the replication of the virus, acting to
 CC cleave its replicative proteins from the polyprotein produced from the
 CC HCV genome. Inhibitors of the two proteins should be effective as
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A
 CC fusion proteins which can be used to identify inhibitors of this type, as
 CC well as enabling structural studies of the protease and
 CC protease:inhibitor complexes. This sequence contains the alpha-helix0-1
 CC variant.
 XX
 SQ Sequence 197 AA;

Query Match 100.0%; Score 1032; DB 21; Length 197;
 Best Local Similarity 100.0%; Pred. No. 1.1e-98;
 Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKKKGSVIVGRINLSGDTAYAAQTGRGEGCGQETSGTGRDNQVEGEVQIVSTAAQTFLA 60
 DB 1 MKKKGSVIVGRINLSGDTAYAAQTGRGEGCGQETSGTGRDNQVEGEVQIVSTAAQTFLA 60
 QY 61 TCINGVCWTVYHGAGTRTIAAPKGPVIOYTNVDKDLVGPAPQGSRLTCTCGSSDLY 120
 DB 61 TCINGVCWTVYHGAGTRTIAAPKGPVIOYTNVDKDLVGPAPQGSRLTCTCGSSDLY 120
 QY 121 LVTRHADVIPVRRGDSRGLSPRPISYLGSGGGLLCPAGHAGVIFRAAVCTRGVAK 180
 DB 121 LVTRHADVIPVRRGDSRGLSPRPISYLGSGGGLLCPAGHAGVIFRAAVCTRGVAK 180
 QY 181 AVDFIPVESLETTMRSP 197
 DB 181 AVDFIPVESLETTMRSP 197

RESULT 2

AA15222
 ID AAB15222 standard; protein; 197 AA.

XX
 AC AAB15222;
 XX
 DT 19-DEC-2000 (first entry)
 XX
 DE Hepatitis C virus NS4A-NS3 fusion protease #4.

XX Hepatitis; NS3 protease; viral replication; chronic liver disease;
 KW liver failure; liver cancer; mutant; mutein.

XX Hepatitis C virus.
 OS Synthetic.

XX WO200040707-A1.

XX 13-JUL-2000.

XX 06-JAN-2000; 2000WO-US00345.

XX 08-JAN-1999; 99US-0115271.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;

XX WPI; 2000-465976/40.

DR N-PSDB; AAA73331.

XX

PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 PT -
 XX
 PS Claim 23; Fig 14; 66pp; English.
 XX
 CC The present sequence is a mutated version of a fusion protein created
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
 CC proteins are both essential for the replication of the virus, acting to
 CC cleave its replicative proteins from the polyprotein produced from the
 CC HCV genome. Inhibitors of the two proteins should be effective as
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A
 CC fusion proteins which can be used to identify inhibitors of this type, as
 CC well as enabling structural studies of the protease and
 CC protease:inhibitor complexes. This sequence contains the alpha-helix0-1
 CC variant.
 XX
 SQ Sequence 197 AA;

Query Match 98.48; Score 1015; DB 21; Length 197;
 Best Local Similarity 98.5%; Pred. No. 6.2e-97;
 Matches 194; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MKKKGSVIVGRINLSGDTAYAAQTGRGEGCGQETSGTGRDNQVEGEVQIVSTAAQTFLA 60
 DB 1 MKKKGSVIVGRINLSGDTAYAAQTGRGEGCGQETSGTGRDNQVEGEVQIVSTAAQTFLA 60
 QY 61 TCINGVCWTVYHGAGTRTIAAPKGPVIOYTNVDKDLVGPAPQGSRLTCTCGSSDLY 120
 DB 61 TCINGVCWTVYHGAGTRTIAAPKGPVIOYTNVDKDLVGPAPQGSRLTCTCGSSDLY 120
 QY 121 LVTRHADVIPVRRGDSRGLSPRPISYLGSGGGLLCPAGHAGVIFRAAVCTRGVAK 180
 DB 121 LVTRHADVIPVRRGDSRGLSPRPISYLGSGGGLLCPAGHAGVIFRAAVCTRGVAK 180
 QY 181 AVDFIPVESLETTMRSP 197
 DB 181 AVDFIPVESLETTMRSP 197

RESULT 3

AA15220

ID AAB15220 standard; protein; 195 AA.

XX
 AC AAB15220;

XX
 DT 19-DEC-2000 (first entry)

XX Hepatitis C virus NS4A-NS3 fusion protease #2.

DE Hepatitis; NS3 protease; viral replication; chronic liver disease;
 KW liver failure; liver cancer; mutant; mutein.

XX Hepatitis C virus.
 OS Synthetic.

XX WO200040707-A1.

XX 13-JUL-2000.

XX 06-JAN-2000; 2000WO-US00345.

XX 08-JAN-1999; 99US-0115271.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;

XX WPI; 2000-465976/40.

DR N-PSDB; AAA73329.

DR

XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
PT amino acid, useful for screening inhibitors that may treat hepatitis C
PT
XX
PS Claim 23: Fig 12: 66pp: English.
XX
CC The present sequence is a mutated version of a fusion protein created
CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
CC proteins are both essential for the replication of the virus, acting to
CC cleave its replicative proteins from the polyprotein produced from the
CC HCV genome. Inhibitors of the two proteins should be effective as
CC antiviral treatments of HCV infection. This is useful as HCV can lead to
CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
CC The present invention concerns a number of NS3 mutants and NS3-NS4A
CC fusion proteins which can be used to identify inhibitors of this type, as
CC well as enabling structural studies of the protease and
CC protease:inhibitor complexes. This sequence contains the alpha-helix0-1
CC variant.
XX
SQ Sequence 195 AA;
Query Match 96.7%; Score 998; DB 21; Length 195;
Best Local Similarity 98.0%; Pred. No. 3.5e-95;
Matches 193; Conservative 1; Mismatches 1; Indels 2; Gaps 1;
QY 1 MKKKGSVIVGRINLSGDTAYAQOTRGECCOETSGTGRDKNOVEGEVQIVSTAAQTFLA 60
DB 1 MKKKGSVIVGRIVLNG--YAAQOTRGECCOETSGTGRDKNOVEGEVQIVSTAAQTFLA 58
QY 61 TCINGVCWTYYHGAGTRTIIASPKGPVIQMYTNVDKDLVGVPAQPGSRSLTPTCTCGSSDLY 120
DB 59 TCINGVCWTYYHGAGTRTIIASPKGPVIQMYTNVDKDLVGVPAQPGSRSLTPTCTCGSSDLY 118
QY 121 LVTRHADVIPVRRGDSRGSLLSPRISYLYKSGGGPLLCPCPAGHAGVIFRAAVCTRGVAK 180
DB 119 LVTRHADVIPVRRGDSRGSLLSPRISYLYKSGGGPLLCPCPAGHAGVIFRAAVCTRGVAK 178
QY 181 AVDFIPVESLETTMRSP 197
DB 179 AVDFIPVESLETTMRSP 195
RESULT 4
AAB15226
ID AAB15226 standard; protein; 197 AA.
XX
AC AAB15226;
XX
DT 19-DEC-2000 (first entry)
XX
DE Hepatitis C virus NS4A-NS3 fusion protease #8.
XX
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW liver failure; liver cancer; mutant; mutein.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
PN WO200040707-A1.
XX
PD 13-JUL-2000.
XX
PF 06-JAN-2000; 2000WO-US00345.
XX
PR 08-JAN-1999; 99US-0115271.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX
DR WPI; 2000-465976/40.

DR N-PSDB; AAB73335.
XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
PT amino acid, useful for screening inhibitors that may treat hepatitis C
PT
XX
PS Example 5; Fig 18: 66pp: English.
XX
CC The present sequence is a mutated version of a fusion protein created
CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
CC proteins are both essential for the replication of the virus, acting to
CC cleave its replicative proteins from the polyprotein produced from the
CC HCV genome. Inhibitors of the two proteins should be effective as
CC antiviral treatments of HCV infection. This is useful as HCV can lead to
CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
CC The present invention concerns a number of NS3 mutants and NS3-NS4A
CC fusion proteins which can be used to identify inhibitors of this type, as
CC well as enabling structural studies of the protease and
CC protease:inhibitor complexes. This sequence contains the alpha-helix0
CC wild-type sequence.
XX
SQ Sequence 197 AA;
Query Match 96.2%; Score 993; DB 21; Length 197;
Best Local Similarity 97.5%; Pred. No. 1.2e-94;
Matches 192; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 MKKKGSVIVGRINLSGDTAYAQOTRGECCOETSGTGRDKNOVEGEVQIVSTAAQTFLA 60
DB 1 MKKKGSVIVGRINLSGDTAYAQOTRGLGCIITSLTGRDKNOVEGEVQIVSTAAQTFLA 60
QY 61 TCINGVCWTYYHGAGTRTIIASPKGPVIQMYTNVDKDLVGVPAQPGSRSLTPTCTCGSSDLY 120
DB 61 TCINGVCWTYYHGAGTRTIIASPKGPVIQMYTNVDKDLVGVPAQPGSRSLTPTCTCGSSDLY 120
QY 121 LVTRHADVIPVRRGDSRGSLLSPRISYLYKSGGGPLLCPCPAGHAGVIFRAAVCTRGVAK 180
DB 121 LVTRHADVIPVRRGDSRGSLLSPRISYLYKSGGGPLLCPCPAGHAGVIFRAAVCTRGVAK 180
QY 181 AVDFIPVESLETTMRSP 197
DB 181 AVDFIPVESLETTMRSP 197
RESULT 5
AAB15223
ID AAB15223 standard; protein; 197 AA.
XX
AC AAB15223;
XX
DT 19-DEC-2000 (first entry)
XX
DE Hepatitis C virus NS4A-NS3 fusion protease #5.
XX
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW liver failure; liver cancer; mutant; mutein.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
PN WO200040707-A1.
XX
PD 13-JUL-2000.
XX
PF 06-JAN-2000; 2000WO-US00345.
XX
PR 08-JAN-1999; 99US-0115271.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX
DR WPI; 2000-465976/40.

DR WPI: 2000-465976/40.
 DR N-PSDB; AAA73332.

XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 PT -

PS Claim 23; Fig 15; 66pp; English.

XX The present sequence is a mutated version of a fusion protein created
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
 CC proteins are both essential for the replication of the virus, acting to
 CC cleave its replicative proteins from the polyprotein produced from the
 CC HCV genome. Inhibitors of the two proteins should be effective as
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A
 CC fusion proteins which can be used to identify inhibitors of this type, as
 CC well as enabling structural studies of the protease and
 CC protease-inhibitor complexes. This sequence contains the alpha-helix0-1
 CC variant.

XX Sequence 197 AA;

Query Match 95.4%; Score 985; DB 21; Length 197;
 Best Local Similarity 97.0%; Pred. No. 8e-94;
 Matches 191; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 MKKGSVIVGRINLSGDTAYAAQTRGEGCQETSQTRDRKNQVEGEVQIVSTAAQTFLA 60
 DB 1 MKKGSVIVGRINLSGDTAYAAQTRGEGCQETSQTRDRKNQVEGEVQIVSTAAQTFLA 60
 QY 61 TCINGVCTVYHGAGTRTIAAPKGPVIOYTNVDKDLVGPAPQGSRLTPTCTGSSDLY 120
 DB 61 TSINGVLTVYHGAGTRTIAAPKGPVTOYTNVDKDLVGPAPQGSRLTPTCTGSSDLY 120
 QY 121 LVTRHADVIPVRRGDSRGLSPRPISYLGSGGGLPCPAGHAGVIFRAAVCTRGVAK 180
 DB 121 LVTRHADVIPVRRGDSRGLSPRPISYLGSGGGLPCPAGHAGVIFRAAVCTRGVAK 180
 QY 181 AVDFIPVESLETTMRSP 197
 DB 181 AVDFIPVESLETTMRSP 197

RESULT 6

AAB15224
 ID AAB15224 standard; protein; 197 AA.

XX AAB15224;

DT 19-DEC-2000 (first entry)

XX Hepatitis C virus NS4A-NS3 fusion protease #6.

XX Hepatitis; NS3 protease; viral replication; chronic liver disease;
 KW liver failure; liver cancer; mutant; mutein.

XX Hepatitis C virus.
 OS Synthetic.

XX WO2000040707-A1.

XX 13-JUL-2000.

XX 06-JAN-2000; 2000WO-US00345.

XX 08-JAN-1999; 99US-0115271.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;

XX

DR WPI: 2000-465976/40.

DR N-PSDB; AAA73333.

XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 PT -

XX Claim 23; Fig 16; 66pp; English.

XX The present sequence is a mutated version of a fusion protein created
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
 CC proteins are both essential for the replication of the virus, acting to
 CC cleave its replicative proteins from the polyprotein produced from the
 CC HCV genome. Inhibitors of the two proteins should be effective as
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A
 CC fusion proteins which can be used to identify inhibitors of this type, as
 CC well as enabling structural studies of the protease and
 CC protease-inhibitor complexes. This sequence contains the alpha-helix0-7
 CC variant.

XX Sequence 197 AA;

Query Match 94.3%; Score 973; DB 21; Length 197;
 Best Local Similarity 95.4%; Pred. No. 1.4e-92;
 Matches 188; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 MKKGSVIVGRINLSGDTAYAAQTRGEGCQETSQTRDRKNQVEGEVQIVSTAAQTFLA 60
 DB 1 MKKGSVIVGRINLSGDTAYAAQTRGEGCQETSQTRDRKNQVEGEVQIVSTAAQTFLA 60
 QY 61 TCINGVCTVYHGAGTRTIAAPKGPVIOYTNVDKDLVGPAPQGSRLTPTCTGSSDLY 120
 DB 61 TSINGVLTVYHGAGTRTIAAPKGPVTOYTNVDKDLVGPAPQGSRLTPTCTGSSDLY 120
 QY 121 LVTRHADVIPVRRGDSRGLSPRPISYLGSGGGLPCPAGHAGVIFRAAVCTRGVAK 180
 DB 121 LVTRHADVIPVRRGDSRGLSPRPISYLGSGGGLPCPAGHAGVIFRAAVCTRGVAK 180
 QY 181 AVDFIPVESLETTMRSP 197
 DB 181 AVDFIPVESLETTMRSP 197

RESULT 7

AAB15225
 ID AAB15225 standard; protein; 197 AA.

XX AAB15225;

XX 19-DEC-2000 (first entry)

XX Hepatitis C virus NS4A-NS3 fusion protease #7.

XX Hepatitis; NS3 protease; viral replication; chronic liver disease;
 KW liver failure; liver cancer; mutant; mutein.

XX Hepatitis C virus.
 OS Synthetic.

XX WO2000040707-A1.

XX 13-JUL-2000.

XX 06-JAN-2000; 2000WO-US00345.

XX 08-JAN-1999; 99US-0115271.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX WPI: 2000-465976/40.
DR N-PSDB: AAA73334.
XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
PT amino acid, useful for screening inhibitors that may treat hepatitis C
PT amino acid, useful for screening inhibitors that may treat hepatitis C
XX
XX Claim 23; Fig 17: 66pp; English.
PS
XX The present sequence is a mutated version of a fusion protein created
CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
CC proteins are both essential for the replication of the virus, acting to
CC cleave its replicative proteins from the polyprotein produced from the
CC HCV genome. Inhibitors of the two proteins should be effective as
CC antiviral treatments of HCV infection. This is useful as HCV can lead to
CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
CC The present invention concerns a number of NS3 mutants and NS3-NS4A
CC fusion proteins which can be used to identify inhibitors of this type, as
CC well as enabling structural studies of the protease and
CC protease-inhibitor complexes. This sequence contains the alpha-helix0-7
XX variant.
XX
SQ Sequence 197 AA;
Query Match 93.3%; Score 963; DB 21; Length 197;
Best Local Similarity 94.9%; Pred. No. 1.5e-91;
Matches 187; Conservative 2; Mismatches 8; Indels 0; Gaps 0;
QY 1 MKKGSVVIVGRINLSGDTAYAAQOTRGEQGCQETSQTGRDKNOVEGEVQIVSTAQTFLA 60
DB 1 MKKGSVVIVGRINLSGDTAYAAQOTRGEQGCQETSQTGRDKNOVEGEVQIVSTAQTFLA 60
QY 61 TCINGVCTVYHGAGTRTIAASPKGPVQMTNVNVDKDLVGVWPAQGSRSLSLTCTCGSSDLY 120
DB 61 TSINGVLTATVYHGAGTRTIAASPKGPVQMTNVNVDKDLVGVWPAQGSRSLSLTCTCGSSDLY 120
QY 121 LVTRHADVIPVRRRGRSGSLSPRPISYLGSGSGPLLCAGHAGVGFRAAVCTRGVAK 180
DB 121 LVTRHADVIPVRRRGRSGSLSPRPISYLGSGSGPLLCAGHAGVGFRAAVCTRGVAK 180
QY 181 AVDFIPVESLETTMRSP 197
DB 181 AVDFIPVESLETTMRSP 197
RESULT 8
AAB15212
ID AAB15212 standard; protein; 195 AA.
XX
AC AAB15212;
XX
DT 19-DEC-2000 (first entry)
DE Hepatitis C virus NS4A-NS3 fusion protease #1.
DE
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW liver failure; liver cancer.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
XX WO2000040707-A1.
PN
XX
PD 13-JUL-2000.
XX
XX 06-JAN-2000; 2000WO-US00345.
PF
XX
PR 08-JAN-1999; 99US-0115271.
XX
XX (BRIM) BRISTOL-MYERS SQUIBB CO.

XX
PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX WPI: 2000-465976/40.
DR N-PSDB: AAA73328.
XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
PT amino acid, useful for screening inhibitors that may treat hepatitis C
PT amino acid, useful for screening inhibitors that may treat hepatitis C
XX
XX Example 2; Fig 10: 66pp; English.
PS
XX The present sequence is a fusion protein created using the Hepatitis C
CC virus (HCV) NS3 and NS4A protease enzymes. These proteins are both
CC essential for the replication of the virus, acting to cleave its
CC replicative proteins from the polyprotein produced from the HCV genome.
CC Inhibitors of the two proteins should be effective as antiviral
CC treatments of HCV infection. This is useful as HCV can lead to chronic
CC liver disease such as cirrhosis, liver failure and liver cancer. The
CC present invention concerns a number of NS3 mutants and NS3-NS4A fusion
CC proteins which can be used to identify inhibitors of this type, as well
CC as enabling structural studies of the protease and protease-inhibitor
CC complexes.
XX
SQ Sequence 195 AA;
Query Match 92.9%; Score 959; DB 21; Length 195;
Best Local Similarity 95.4%; Pred. No. 3.9e-91;
Matches 188; Conservative 1; Mismatches 6; Indels 2; Gaps 1;
QY 1 MKKGSVVIVGRINLSGDTAYAAQOTRGEQGCQETSQTGRDKNOVEGEVQIVSTAQTFLA 60
DB 1 MKKGSVVIVGRIVLNG--AYAQOTRGLGCIITSLTRDKNOVEGEVQIVSTAQTFLA 58
QY 61 TCINGVCTVYHGAGTRTIAASPKGPVQMTNVNVDKDLVGVWPAQGSRSLSLTCTCGSSDLY 120
DB 59 TCINGVCTVYHGAGTRTIAASPKGPVQMTNVNVDKDLVGVWPAQGSRSLSLTCTCGSSDLY 118
QY 121 LVTRHADVIPVRRRGRSGSLSPRPISYLGSGSGPLLCAGHAGVGFRAAVCTRGVAK 180
DB 119 LVTRHADVIPVRRRGRSGSLSPRPISYLGSGSGPLLCAGHAGVGFRAAVCTRGVAK 178
QY 181 AVDFIPVESLETTMRSP 197
DB 179 AVDFIPVESLETTMRSP 195
RESULT 9
AAR40120
ID AAR40120 standard; protein; 3011 AA.
XX
AC AAR40120;
XX
DT 25-MAR-2003 (updated)
DT 27-JAN-1994 (first entry)
XX
DE HCV genomic amino acid sequence isolated from infected human LG.
XX
KW Hepatitis C Virus; Non-A, non-B hepatitis virus; HCV; NANBHV;
KW human growth hormone; HGH; secretion signal; fusion protein;
KW vaccine.
XX
OS Hepatitis C Virus.
XX
PN WO9315193-A1.
XX
PD 05-AUG-1993.
XX
PF 29-JAN-1993; 93WO-US00907.
XX
XX 31-JAN-1992; 92US-0830024.
XX

PA (ABBO) ABBOTT LAB.
 XX Bode SL, Casey JM, Desal SM, Devare SG, Frail DE;
 PI Yamaguchi J, Zeck BJ;
 XX WPI; 1993-258673/32.
 XX New plasmid pHCV-162 is a mammalian expression systems for HCV
 PT proteins - useful for diagnosing HCV infection and as vaccines
 PT for preventing HCV infection
 XX
 XX Example 1; Page 39-49; 100pp; English.
 XX RNA was isolated from the plasma of a HCV seropositive human
 CC (designated "LG") and cDNA was prepared from it. The cDNA was
 CC PCR amplified using specific primers with sequences based
 CC on the prototype HCV-1 cDNA sequence (GENBANK M62321). Further
 CC amplification using nested primers resulted in 7 adjacent HCV DNA
 CC fragments which could be assembled into a full-length sequence. The
 CC DNA sequence was determined and translated into the genomic amino
 CC acid sequence. Comparison of the LG genomic amino acid sequence
 CC with that from HCV-1 showed 134 amino acid differences.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 3011 AA;

Query Match 87.5%; Score 902.5; DB 14; Length 3011;
 Best Local Similarity 85.8%; Pred. No. 1.1e-83;
 Matches 175; Conservative 9; Mismatches 11; Indels 9; Gaps 1;
 QY 3 KKGSVIVGRIN-----LSGDTAYAOQTREEGCQETSQTRDKNOVEGEVQIVST 53
 DB 1005 RRGREILLGPADGMVSKGWRLIAPITAYAOQTRGLLCIITSLTGRDKNOVEGEVQIVST 1064
 QY 54 AAQTFLATCINGCVTVYHGAGTRTIASPKGPVIQMTYNDKDLVGPAPQGSRLTPCT 113
 DB 1065 AAQTFLATCINGCVTVYHGAGTRTIASPKGPVIQMTYNDKDLVGPAPQGSRLTPCT 1124
 QY 114 CGSSDLXLVTRHADYIPVRRRGDSRGLSPRPISYLYKSGSGGPLLCPAGHAVGIFRAAV 173
 DB 1125 CGSSDLXLVTRHADYIPVRRRGDSRGLSPRPISYLYKSGSGGPLLCPAGHAVGIFRAAV 1184
 QY 174 CTRGVAKAVDFIPVESLETTMRSP 197
 DB 1185 CTRGVAKAVDFIPVESLETTMRSP 1208

RESULT 10
 AAP92041
 ID AAP92041 standard; protein; 1766 AA.
 AC AAP92041;
 XX 25-MAR-2003 (updated)
 DT 02-MAR-1990 (first entry)
 XX
 XX Sequence encoded in the hepatitis C virus (HCV) cDNA inserts in clones
 DE 141, 11b, 7f, 7e, 8h, 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c, 14c, 8f,
 DE 33f, 33g and 39c.
 XX
 XX Hepatitis C virus (HCV); non-A, non-B hepatitis (HANBH)
 KW Hepatitis C virus.
 OS Hepatitis C virus.
 XX EP318216-A.
 XX 31-MAY-1989.
 PD
 XX 18-NOV-1988; 88EP-0310922.
 PF
 XX 18-NOV-1987; 87US-0122714.
 PR 30-DEC-1987; 87US-0139886.
 PR 26-FEB-1988; 88US-0161072.
 PR

PR 06-MAY-1988; 88US-0191263.
 PR 26-OCT-1988; 88US-0263584.
 PR 14-NOV-1988; 88US-0271450.
 XX (CHIR) CHIRON CORP.
 XX Houghton M, Choo QL, Kuo G;
 XX WPI; 1989-159274/22.
 DR N-PSDB; AAN92097.
 XX
 XX Purified hepatitis C virus
 PT - and associated nucleic acids and polypeptide(s)
 XX
 XX Claim 13; Figure 26-1, 26-2, 26-3, 26-4, 26-5, 26-6; 139pp; English.
 XX It is the sequence encoded in the open reading frame of hepatitis C virus
 CC cDNA inserts in clones 141, 11b, 7f, 7e, 8h, 33c, 40b, 37b, 35, 36,
 CC 81, 32, 33b, 25c, 14c, 8f, a33f, 33g and 39c. It is antigenic and could
 CC be used in immunoassay reagents and vaccines and to generate antibodies
 CC useful in diagnosis and passive immunotherapy for HCV infection/non-A.
 CC non-B hepatitis.
 CC (Updated on 25-MAR-2003 to correct PR field.)
 CC (Updated on 25-MAR-2003 to correct PI field.)
 XX
 SQ Sequence 1766 AA;

Query Match 87.4%; Score 901.5; DB 10; Length 1766;
 Best Local Similarity 85.8%; Pred. No. 6.9e-84;
 Matches 175; Conservative 9; Mismatches 11; Indels 9; Gaps 1;
 QY 3 KKGSVIVGRIN-----LSGDTAYAOQTREEGCQETSQTRDKNOVEGEVQIVST 53
 DB 289 RRGREILLGPADGMVSKGWRLIAPITAYAOQTRGLLCIITSLTGRDKNOVEGEVQIVST 348
 QY 54 AAQTFLATCINGCVTVYHGAGTRTIASPKGPVIQMTYNDKDLVGPAPQGSRLTPCT 113
 DB 349 AAQTFLATCINGCVTVYHGAGTRTIASPKGPVIQMTYNDKDLVGPAPQGSRLTPCT 408
 QY 114 CGSSDLXLVTRHADYIPVRRRGDSRGLSPRPISYLYKSGSGGPLLCPAGHAVGIFRAAV 173
 DB 409 CGSSDLXLVTRHADYIPVRRRGDSRGLSPRPISYLYKSGSGGPLLCPAGHAVGIFRAAV 468
 QY 174 CTRGVAKAVDFIPVESLETTMRSP 197
 DB 469 CTRGVAKAVDFIPVESLETTMRSP 492

RESULT 11
 AAP90158
 ID AAP90158 standard; protein; 1786 AA.
 AC AAP90158;
 XX 25-MAR-2003 (updated)
 DT 10-NOV-1989 (first entry)
 XX
 XX Protein sequence of hepatitis c virus composite cDNA.
 DE
 XX Hepatitis C virus; vaccine.
 KW Pan troglodytes.
 OS
 XX GB2212511-A.
 XX 26-JUL-1989.
 PD
 XX 18-NOV-1988; 88GB-0027024.
 PF
 XX 18-NOV-1987; 87US-0122714.
 PR 30-DEC-1987; 87US-0139886.
 PR 26-FEB-1988; 88US-0161072.
 PR 26-OCT-1988; 88US-0263584.
 PR

XX (CHIR) CHIRON CORPORATION.
 XX Houghton M, Choo QL, Kuo G;
 XX WPI: 1989-215054/30.
 XX N-PSDB; AAN90327.
 XX Hepatitis C virus gene - used for prodn. of polynucleotide probes,
 PT polypeptide(s) and antibodies for diagnosis, prevention and treatment
 PT of infection.
 XX Disclosure; fig 26; 30pp; English.
 XX The sequence is encoded by the composite cDNA of AAN90327. These
 CC antigens react with antibodies in patients with non-A non-B hepatitis
 CC (NANBH). They can be used to diagnose HCV-induced NANBH, to raise
 CC antibodies for immunoassay or treatment, or to produce vaccines.
 CC (Updated on 25-MAR-2003 to correct PR field.)
 XX Sequence 1786 AA;
 SQ
 Query Match 87.4%; Score 901.5; DB 10; Length 1786;
 Best Local Similarity 85.8%; Pred. No. 7e-84;
 Matches 175; Conservative 9; Mismatches 11; Indels 9; Gaps 1;
 QY 3 KGSVVIVGRIN-----LSGDTAYAAQTRGEGCGQETSGTRDKNOVEGEVQIVST 53
 Db 289 RRGREILLGPADGMVSKGWRLLAPITAYAAQTRGLLCIITSITGRDKNOVEGEVQIVST 348
 QY 54 AAQTFLATCINGVCWTVYHGAGTRTIASPKGPVIQMYTNVDKDLVGPAPQGSRLTPTCT 113
 Db 349 AAQTFLATCINGVCWTVYHGAGTRTIASPKGPVIQMYTNVDQDLVGPAPQGSRLTPTCT 408
 QY 114 CGSSDLVLTTRHADVTPVRRGRDGRGSLSPRISYLVKSGSGGPLLCPCAGHAGVIFRAAV 173
 Db 409 CGSSDLVLTTRHADVTPVRRGRDGRGSLSPRISYLVKSGSGGPLLCPCAGHAGVIFRAAV 468
 QY 174 CTRGVAKAVDFIPVESLETTMRSP 197
 Db 469 CTRGVAKAVDFIPVENLETTMRSP 492
 RESULT 12
 AAP90164
 ID AAP90164 standard; protein; 2261 AA.
 XX AAP90164;
 AC
 XX 25-MAR-2003 (updated)
 DT 01-NOV-1989 (first entry)
 XX
 XX Peptide encoded by composite hepatitis C virus cDNA.
 DE
 XX Hepatitis C virus; clone 12f; clone 15e; probe; vaccine.
 KW
 XX Pan troglodytes.
 OS
 XX GB2212511-A.
 PN
 XX 26-JUL-1989.
 PD
 XX 18-NOV-1988; 88GB-0027024.
 XX
 XX 18-NOV-1987; 87US-0122714.
 XX 30-DEC-1987; 87US-0139886.
 XX 26-FEB-1988; 88US-0161072.
 XX 06-MAY-1988; 88US-0191263.
 XX 26-OCT-1988; 88US-0263584.
 XX 14-NOV-1988; 88US-0271450.
 XX
 XX (CHIR) CHIRON CORPORATION.
 PA
 XX Houghton M, Choo QL, Kuo G;
 XX WPI: 1989-159274/22.
 XX N-PSDB; AAN92103.
 DR

DR WPI: 1989-215054/30.
 DR N-PSDB; AAN90331.
 XX Hepatitis C virus gene - used for prodn. of polynucleotide probes,
 PT polypeptide(s) and antibodies for diagnosis, prevention and
 PT treatment of infection.
 XX Disclosure; fig 32; 235pp; English.
 XX The sequence is the peptide encoded by the composite hepatitis C
 CC virus (HCV) cDNA of AAN90331. The polypeptides are used to diagnose
 CC HCV-induced NANBH, to raise antibodies for immunoassay or treatment,
 CC or to produce vaccines.
 CC (Updated on 25-MAR-2003 to correct PR field.)
 XX Sequence 2261 AA;
 SQ
 Query Match 87.4%; Score 901.5; DB 10; Length 2261;
 Best Local Similarity 85.8%; Pred. No. 9.6e-84;
 Matches 175; Conservative 9; Mismatches 11; Indels 9; Gaps 1;
 QY 3 KGSVVIVGRIN-----LSGDTAYAAQTRGEGCGQETSGTRDKNOVEGEVQIVST 53
 Db 380 RRGREILLGPADGMVSKGWRLLAPITAYAAQTRGLLCIITSITGRDKNOVEGEVQIVST 439
 QY 54 AAQTFLATCINGVCWTVYHGAGTRTIASPKGPVIQMYTNVDKDLVGPAPQGSRLTPTCT 113
 Db 440 AAQTFLATCINGVCWTVYHGAGTRTIASPKGPVIQMYTNVDQDLVGPAPQGSRLTPTCT 499
 QY 114 CGSSDLVLTTRHADVTPVRRGRDGRGSLSPRISYLVKSGSGGPLLCPCAGHAGVIFRAAV 173
 Db 500 CGSSDLVLTTRHADVTPVRRGRDGRGSLSPRISYLVKSGSGGPLLCPCAGHAGVIFRAAV 559
 QY 174 CTRGVAKAVDFIPVESLETTMRSP 197
 Db 560 CTRGVAKAVDFIPVENLETTMRSP 583
 RESULT 13
 AAP92047
 ID AAP92047 standard; protein; 2301 AA.
 XX AAP92047;
 AC
 XX 25-MAR-2003 (updated)
 DT 02-MAR-1990 (first entry)
 XX
 XX Sequence encoded in the hepatitis C virus (HCV) cDNA inserts in clones
 DE 12f through 15e.
 DE
 XX Hepatitis C virus (HCV); non-A, non-B hepatitis (NANBH).
 KW
 XX Hepatitis C virus.
 OS
 XX EP318216-A.
 PN
 XX 31-MAY-1989.
 PD
 XX 18-NOV-1988; 88EP-0310922.
 XX
 XX 18-NOV-1987; 87US-0122714.
 XX 30-DEC-1987; 87US-0139886.
 XX 26-FEB-1988; 88US-0161072.
 XX 06-MAY-1988; 88US-0191263.
 XX 26-OCT-1988; 88US-0263584.
 XX 14-NOV-1988; 88US-0271450.
 XX
 XX (CHIR) CHIRON CORP.
 PA
 XX Houghton M, Choo QL, Kuo G;
 XX WPI: 1989-159274/22.
 XX N-PSDB; AAN92103.
 DR

XX Purified hepatitis C virus
PT - and associated nucleic acids and polypeptide(s)
XX Claim 13: Figure 32-1 - 32-7; 139 pp; English.
XX It is the sequence encoded in the open reading frame of hepatitis C virus
CC (HCV) cDNA inserts in clones 12f through 15e. It is antigenic and could
CC be used in immunoassay reagents and vaccines and to generate antibodies
CC useful in diagnosis and passive immunotherapy for HCV infection/non-A,
CC non-B hepatitis.
CC (Updated on 25-MAR-2003 to correct PR field.)
CC (Updated on 25-MAR-2003 to correct PI field.)
XX Sequence 2301 AA;
SQ
Query Match 87.4%; Score 901.5; DB 10; Length 2301;
Best Local Similarity 85.8%; Pred. No. 9.9e-84;
Matches 175; Conservative 9; Mismatches 11; Indels 9; Gaps 1;
QY 3 KKGSVVIVGRIN-----LSGDTAYAOQTRGEGCOETSQTGRDNQVEGEVQIVST 53
DB 380 RRGREILLGPADGMVSKGWRLLAPITAYAOQTRGLGCIITSLTGRDNQVEGEVQIVST 439
QY 54 AAQTFELATCINGCVTVYHGAGTRTIAASPKGPVIOMYTNVDKDLVGPAPOGSRSLSPTCT 113
DB 440 AAQTFELATCINGCVTVYHGAGTRTIAASPKGPVIOMYTNVDKDLVGPAPOGSRSLSPTCT 499
QY 114 CGSSDLYLVTRHADYIPVRRRGDSRGLSPRPISYLGSSGGPLLCPCAGHAVGIFRAAV 173
DB 500 CGSSDLYLVTRHADYIPVRRRGDSRGLSPRPISYLGSSGGPLLCPCAGHAVGIFRAAV 559
QY 174 CTRGVAKAVDFIPVESLETMRSP 197
DB 560 CTRGVAKAVDFIPVENLETMRSP 583

RESULT 14
AAP92050
ID AAP92050 standard; protein: 2436 AA.
XX
AC AAP92050;
XX 25-MAR-2003 (updated)
DT 02-MAR-1990 (first entry)
XX
XX Sequence encoded in the hepatitis C virus (HCV) cDNA inserts in clones
DE K9-1 through 15e.
XX Hepatitis C virus (HCV); non-A, non-B hepatitis (HANBH)
XX Hepatitis C virus.
OS
XX EP318216-A.
XX
XX 31-MAY-1989.
XX
XX 18-NOV-1988; 88EP-0310922.
XX
XX 18-NOV-1987; 87US-0122714.
XX 30-DEC-1987; 87US-0139886.
XX 26-FEB-1988; 88US-0161072.
XX 06-MAY-1988; 88US-0191263.
XX 26-OCT-1988; 88US-0263584.
XX 14-NOV-1988; 88US-0271450.
XX
XX (CHIR) CHIRON CORP.
XX
XX Houghton M, Choo QL, Kuo G;
PI WPI; 1989-159274/22.
XX
XX N-PSDB; AAP92106.
XX

PT Purified hepatitis C virus
PT - and associated nucleic acids and polypeptide(s)
XX Claim 13: Figure 47-1 - 47-8; 139 pp; English.
XX It is the sequence encoded in the open reading frame of hepatitis C virus
CC (HCV) cDNA inserts in clones K9-1 through 15e. It is antigenic and could
CC be used in immunoassay reagents and vaccines and to generate antibodies
CC useful in diagnosis and passive immunotherapy for HCV infection/non-A,
CC non-B hepatitis.
CC (Updated on 25-MAR-2003 to correct PR field.)
CC (Updated on 25-MAR-2003 to correct PI field.)
XX Sequence 2436 AA;
SQ
Query Match 87.4%; Score 901.5; DB 10; Length 2436;
Best Local Similarity 85.8%; Pred. No. 1.1e-83;
Matches 175; Conservative 9; Mismatches 11; Indels 9; Gaps 1;
QY 3 KKGSVVIVGRIN-----LSGDTAYAOQTRGEGCOETSQTGRDNQVEGEVQIVST 53
DB 555 RRGREILLGPADGMVSKGWRLLAPITAYAOQTRGLGCIITSLTGRDNQVEGEVQIVST 614
QY 54 AAQTFELATCINGCVTVYHGAGTRTIAASPKGPVIOMYTNVDKDLVGPAPOGSRSLSPTCT 113
DB 615 AAQTFELATCINGCVTVYHGAGTRTIAASPKGPVIOMYTNVDKDLVGPAPOGSRSLSPTCT 674
QY 114 CGSSDLYLVTRHADYIPVRRRGDSRGLSPRPISYLGSSGGPLLCPCAGHAVGIFRAAV 173
DB 675 CGSSDLYLVTRHADYIPVRRRGDSRGLSPRPISYLGSSGGPLLCPCAGHAVGIFRAAV 734
QY 174 CTRGVAKAVDFIPVESLETMRSP 197
DB 735 CTRGVAKAVDFIPVENLETMRSP 758

RESULT 15
AAP90288
ID AAP90288 standard; protein: 2436 AA.
XX
AC AAP90288;
XX 25-MAR-2003 (updated)
DT 19-JUL-2001 (updated)
DT 01-NOV-1989 (first entry)
XX
XX Peptide encoded by composite hepatitis C cDNA.
XX Hepatitis C virus; clone 15e; clone K9-1; probe; vaccine.
XX Pan troglodytes.
OS
XX GB2212511-A.
XX
XX 26-JUL-1989.
XX
XX 18-NOV-1988; 88GB-0027024.
XX
XX 18-NOV-1987; 87US-0122714.
XX 30-DEC-1987; 87US-0139886.
XX 26-FEB-1988; 88US-0161072.
XX 26-OCT-1988; 88US-0263584.
XX
XX (CHIR) CHIRON CORPORATION.
XX
XX Houghton M, Choo QL, Kuo G;
PI WPI; 1989-215054/30.
XX
XX N-PSDB; AAP90336.
XX
XX Hepatitis C virus gene - used for prodn. of polynucleotide probes,
XX polypeptide(s) and antibodies for diagnosis, prevention and
XX treatment of infection.
XX

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 19:02:22 ; Search time 16.2134 Seconds
(without alignments)
1168.492 Million cell updates/sec

Title: US-09-965-594-14

Perfect score: 1032
Sequence: 1 MKKKGWVIVGRINLSGDTA.....VAKAVDFIPVESLETTMRSP 197

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96169682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	901.5	87.4	3011	1 GNMVC3	genome polyprotein
2	900.5	87.3	3011	1 S40770	genome polyprotein
3	896.5	85.9	3011	1 GNMVCH	genome polyprotein
4	894.5	82.8	3010	1 GNMVTW	genome polyprotein
5	890.5	82.4	3010	1 GNMVTC	genome polyprotein
6	850.5	82.4	3010	1 GNMVVC	genome polyprotein
7	844.5	81.8	3010	1 A45573	genome polyprotein
8	828.5	80.3	3010	1 S18030	genome polyprotein
9	766.5	74.3	3014	1 JC5620	genome polyprotein
10	674	65.3	3033	1 JQ1303	genome polyprotein
11	672	65.1	3033	1 GNMVJ8	genome polyprotein
12	257	24.9	3005	2 T08841	polyprotein - dour
13	251	24.3	2970	2 T08839	polyprotein - marm
14	85.5	8.3	495	2 E71360	hypothetical prote
15	82	7.9	452	2 I39383	angio-associated m
16	81	7.8	477	2 E75392	hypothetical prote
17	80.5	7.8	590	2 E81104	nitrate/nitrite se
18	80.5	7.8	590	2 C81911	nitrate/nitrite se
19	78.5	7.6	209	2 H83144	probable ornatic
20	78.5	7.6	981	2 T18234	beta transducin ho
21	78	7.6	915	2 F81196	transferrin-bindin
22	77.5	7.5	398	2 B71284	probable periplasm
23	77.5	7.5	2663	1 S28261	centromere protein
24	77	7.5	354	2 T49806	hypothetical prote
25	77	7.5	3739	2 T17410	polyketide synthas
26	76.5	7.4	270	2 T06118	hypothetical prote
27	76.5	7.4	492	2 AH1030	probable exported
28	76	7.4	140	2 C72705	hypothetical prote
29	76	7.4	583	2 S38789	ferredoxin-nitrite

30	76	7.4	911	2 JN0821	transferrin-bindin
31	76	7.4	2508	2 S61441	surface-associated
32	75.5	7.3	329	2 C96033	probable regulator
33	75.5	7.3	1334	2 AB1775	hypothetical prote
34	75	7.3	322	2 D87603	glycosyl transfera
35	75	7.3	603	1 VCFVER	env polyprotein -
36	75	7.3	910	2 C81832	transferrin-bindin
37	75	7.3	1293	2 T30871	orsellinic acid sy
38	74.5	7.2	415	2 S70401	zona pellucida gly
39	74.5	7.2	424	2 AH2406	hypothetical prote
40	74.5	7.2	755	2 S23441	hypothetical prote
41	74.5	7.2	846	2 T04533	hypothetical prote
42	74	7.2	456	2 S30922	ferredoxin-nitrite
43	74	7.2	1165	1 GNLJGL	HIV-1 retropepsin
44	73.5	7.1	239	2 G87265	conserved hypothet
45	73.5	7.1	317	2 S76618	hypothetical prote

ALIGNMENTS

RESULT 1

GNMVC3

genome polyprotein - hepatitis C virus (strain HCV-1)

N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain HCV-1) (nonstr protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 19-Jan-2001

C:Accession: A39166; PQ0403; PQ0404

R:Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.;

Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991

A:Title: Genetic organization and diversity of the hepatitis C virus.

A:Reference number: A39166; MUID:91172826; PMID:1848704

A:Accession: A39166

A:Molecule type: mRNA

A:Residues: 1-3011 <CHO>

A:Cross-references: GB:M62321; NID:g329873; PIDN:AAA45676.1; PID:g329874

J. Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap,

R. Gen. Virol. 73, 1131-1141, 1992

A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship

A:Reference number: PQ0393; MUID:92268871; PMID:1316939

A:Accession: PQ0403

A:Molecule type: genomic RNA

A:Residues: 1577-1633 <CHA>

A:Cross-references: DDBJ:D10128

A:Experimental source: Isolates E-b16

A:Accession: PQ0404

A>Status: preliminary

A:Molecule type: genomic RNA

A:Residues: 1577-1633 <CH2>

A:Experimental source: Isolates E-b17

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruc

F:115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: envelope protein M #status predicted <EPM>

F:192-389/Product: major envelope protein E #status predicted <NEE>

F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F:1007-1615/Product: hepatitis C virus genome polyprotein

F:1230-1237/Region: nucleotide-binding motif A (P-loop)

F:1312-1317/Region: nucleotide-binding motif B

F:1316-1319/Region: DEX motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>

F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,207

Query Match 87.4%; Score 901.5; DB 1; Length 3011;

Best Local Similarity 85.8%; Pred. No. 3.7e-75;

Matches 175; Conservative 9; Mismatches 11; Indels 9; Gaps 1;

Qy 3 KKGWVIVGRIN-----LSGDTAYAOQTRGEGCOETSGTRKNQVGEVIVST 53

::: ::::: : : ||||| || || ||||| ||||| |||||


```

RESULT 3
GNWVCH
genome polyprotein - hepatitis C virus (strain H)
N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstruc
Protein NS4; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
A:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C:Accession: A36814; A41546
R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
submitted to GenBank July 1992
A:Description: Genomic structure of the human prototype strain H of hepatitis C virus
A:Reference number: A36814
A:Accession: A36814
A:Molecule type: genomic RNA
A:Residues: 1-3011 <INC>
A:Cross-references: GB:M67463; NID:g329737; PIDN:AAA45534.1; PID:g329738
R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
A:title: Genomic structure of the human prototype strain H of hepatitis C virus: comp
A:Reference number: A41546; MOID:92052256; PMID:1658600
A:Contents: annotation
A:Note: neither amino acid nor nucleotide sequence is given
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruct
F:1-115/Product: capsid protein C #status predicted <GPC>
F:116-191/Product: envelope protein M #status predicted <EPH>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepacivirin #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NA4>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NA4B>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F:136,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240
Query Match      85.9%   Score 886.5; DB 1; Length 3011;
Best Local Similarity 83.8%; Pred. No. 9,2e-74;
Matches 171; Conservative 10; Mismatches 14; Indels 9; Gaps 1;

Qy    3 KKGSWTVGRIN-----LSGDTAYAOOTRGEGCGTETSOTGRDKNOVEGEVQIVST 53
       ::::: :|: |||||::| |:|||||::| |:|||||::| |:|||||::|
Db    1005 RRGQEILLGPADGMVSGKWRLLAPITAYAQTGRGLLCIIITSLTGRDKNOVEGEVQIVST 1064

Qy    54 AAQTFLATCLNGVCWTYYHGAGRTIASPKPGVPVQMVTYNDVKDLVGWPAPQGSRSLTPCT 113
       |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db    1065 ATQTFLATCLNGVCWTYYHGAGRTIASPKGPVIQTTNVDDQDVLGWPAFGSRLTPCT 1124

Qy    114 CGSSDLYLVTRHADVIPRRGGRSGLLSPRPISYLKSGSGGPLCCPAGHAVGIFRAAV 173
       |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db    1125 CGSSDLYLVTRHADVIPRRGGRSGLLSPRPISYLKSGSGGPLLCPGTGHAVGLFAAA 1184

Qy    174 CTGCAKAVDFIPVESLETHMRSP 197
       |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db    1185 CTGCAKAVDFIPVENLETTHMRSP 1208


RESULT 4
GNWVTH
genome polyprotein - hepatitis C virus (strain Taiwan)
N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstruc
Protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
A:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C:Accession: A40244
R:Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.
Virology 188, 102-113, 1992

```

[illegible]

Db 1065 ATQSFATCVNGVWTVYHAGSGSKTLGAPGPIITQMYTINVDQDLVGPAPPGARSMPCT 1124
Qy 114 CGSSDLVLTTRHADVPVRRGDSRGLSPRPISYLGKSGGGLPCPAGHAVGIFRAAV 173
Db 1125 CGSSDLVLTTRHADVPVRRGDSRGLSPRPISYLGKSGGGLPCPAGHAVGIFRAAV 1184
Qy 174 CTRGVAKAVDFIPVESLETTMRSP 197
Db 1185 CTRGVAKAVDFIPVESMETTRSP 1208

RESULT 7
A:Variety: isolate JK1
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
A:Accession: S18030
R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
Submitted to the EMBL data library, September 1991
A:Description: A whole genome of hepatitis C virus cDNA was isolated from a single pa
A:Reference number: S18028
A:Accession: S18030
A:Molecule type: genomic RNA
A:Residues: 1-3010 <HON>
A:Cross-references: EMBL:X61596; NID:g59478; PIDN:CAA43793.1; PID:g59479
R:Honda, M.; Kaneko, S.; Unoura, M.; Kobayashi, K.; Murakami, S.
Arch. Virol. 128, 163-169, 1993
A:Title: Sequence analysis of putative structural regions of hepatitis C virus isolat
A:Reference number: A45573; MUID:92295714; PMID:1318627
A:Accession: A45573
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-3010 <GAN>
A:Cross-references: GB:D01171; NID:g221612; PIDN:BAA01943.1; PID:g221613
A:Experimental source: HCV-JT
A:Note: sequence extracted from NCBI backbone (NCBIN:106206, NCBIP:106207)
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPW>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B

Query Match 81.8%; Score 844.5; DB 1; Length 3010;
Best Local Similarity 77.5%; Pred. No. 7.4e-70;
Matches 158; Conservative 20; Mismatches 17; Indels 9; Gaps 1;
Qy 3 KGSVVIVGRIN-----LSGDTAYAAQTGREGGCGTSTQGRDKNOVEGEVQIVST 53
Db 1005 RRGREILLGPADSGEQWRLAPITAYAAQTGRLGCVITSLTGRDKNOVEGEVQIVST 1064
Qy 54 AAQTFLATCINGVWTVYHAGSRTTASPKGPIQMYTINVDKLVGWPAPGSRSLTPCT 113
Db 1065 ATQSFATCVNGVWTVYHAGSRTTASPKGPIQMYTINVDQDLVGHAPGARSLETPCT 1124
Qy 114 CGSSDLVLTTRHADVPVRRGDSRGLSPRPISYLGKSGGGLPCPAGHAVGIFRAAV 173
Db 1125 CGSSDLVLTTRHADVPVRRGDSRGLSPRPISYLGKSGGGLPCPAGHAVGIFRAAV 1184
Qy 174 CTRGVAKAVDFIPVESLETTMRSP 197
Db 1185 CTRGVAKAVDFIPVESMETTRSP 1208

genome polyprotein - hepatitis C virus (isolate JK1)
N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
A:Accession: A45573
R:Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata,
Virus Res. 23, 39-53, 1992
A:Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: s
A:Reference number: A45573; MUID:92295714; PMID:1318627
A:Accession: A45573
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-3010 <GAN>
A:Cross-references: GB:D01171; NID:g221612; PIDN:BAA01943.1; PID:g221613
A:Experimental source: HCV-JT
A:Note: sequence extracted from NCBI backbone (NCBIN:106206, NCBIP:106207)
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPW>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B

Query Match 80.3%; Score 828.5; DB 1; Length 3010;
Best Local Similarity 76.5%; Pred. No. 2.3e-68;
Matches 156; Conservative 20; Mismatches 19; Indels 9; Gaps 1;
Qy 3 KGSVVIVGRIN-----LSGDTAYAAQTGREGGCGTSTQGRDKNOVEGEVQIVST 53
Db 1005 RRGREILLGPADSGEQWRLAPITAYAAQTGRLGCVITSLTGRDKNOVEGEVQIVST 1064
Qy 54 AAQTFLATCINGVWTVYHAGSRTTASPKGPIQMYTINVDKLVGWPAPGSRSLTPCT 113
Db 1065 ATQSFATCVNGVWTVYHAGSRTTASPKGPIQMYTINVDQDLVGHAPGARSLETPCT 1124
Qy 114 CGSSDLVLTTRHADVPVRRGDSRGLSPRPISYLGKSGGGLPCPAGHAVGIFRAAV 173
Db 1125 CGSSDLVLTTRHADVPVRRGDSRGLSPRPISYLGKSGGGLPCPAGHAVGIFRAAV 1184
Qy 174 CTRGVAKAVDFIPVESLETTMRSP 197
Db 1185 CTRGVAKAVDFIPVESMETTRSP 1208

genome polyprotein - hepatitis C virus (isolate EUH1480)
N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstru
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
A:Accession: JC5620
R:Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.
Biochem. Biophys. Res. Commun. 236, 44-49, 1997
A:Title: The complete coding sequence of hepatitis C virus genotype 5a, the predomina
A:Reference number: JC5620; MUID:97366593; PMID:9223423

RESULT 9
JC5620
genome polyprotein - hepatitis C virus (isolate EUH1480)
N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstru
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
A:Accession: JC5620
R:Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.
Biochem. Biophys. Res. Commun. 236, 44-49, 1997
A:Title: The complete coding sequence of hepatitis C virus genotype 5a, the predomina
A:Reference number: JC5620; MUID:97366593; PMID:9223423

A:Variety: isolate JK1
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 23-Mar-2001
A:Accession: S18030; S18032; A48332; S18029
R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
Submitted to the EMBL data library, September 1991
A:Description: A whole genome of hepatitis C virus cDNA was isolated from a single pa
A:Reference number: S18028
A:Accession: S18030
A:Molecule type: genomic RNA
A:Residues: 1-3010 <HON>
A:Cross-references: EMBL:X61596; NID:g59478; PIDN:CAA43793.1; PID:g59479
R:Honda, M.; Kaneko, S.; Unoura, M.; Kobayashi, K.; Murakami, S.
Arch. Virol. 128, 163-169, 1993
A:Title: Sequence analysis of putative structural regions of hepatitis C virus isolat
A:Reference number: A48332; MUID:93119270; PMID:8380322
A:Accession: S18030
A:Molecule type: genomic RNA
A:Residues: 1-547, 'T', 549-621, 'V', 623-624, 'S', 626-652, 'DL', 655-761, 'T', 763-782 <HOW>
A:Cross-references: EMBL:X61591
A:Note: this sequence is inconsistent with the nucleotide translation
as Trp, and TTC for residue 771 as Ser
A:Note: sequence extracted from NCBI backbone (NCBIN:121747, NCBIP:121748)
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; se
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPW>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,250,305,417,423,448,532,540,556,562,623,645/Binding site: carbohydrate

Query Match 80.3%; Score 828.5; DB 1; Length 3010;
Best Local Similarity 76.5%; Pred. No. 2.3e-68;
Matches 156; Conservative 20; Mismatches 19; Indels 9; Gaps 1;
Qy 3 KGSVVIVGRIN-----LSGDTAYAAQTGREGGCGTSTQGRDKNOVEGEVQIVST 53
Db 1005 RRGREILLGPADSGEQWRLAPITAYAAQTGRLGCVITSLTGRDKNOVEGEVQIVST 1064
Qy 54 AAQTFLATCINGVWTVYHAGSRTTASPKGPIQMYTINVDKLVGWPAPGSRSLTPCT 113
Db 1065 ATQSFATCVNGVWTVYHAGSRTTASPKGPIQMYTINVDQDLVGHAPGARSLETPCT 1124
Qy 114 CGSSDLVLTTRHADVPVRRGDSRGLSPRPISYLGKSGGGLPCPAGHAVGIFRAAV 173
Db 1125 YGSSDLVLTTRHADVPVRRGDSRGLSPRPISYLGKSGGGLPCPAGHAVGIFRAAV 1184
Qy 174 CTRGVAKAVDFIPVESLETTMRSP 197
Db 1185 CTRGVAKAVDFIPVESMETTRSP 1208

genome polyprotein - hepatitis C virus (isolate EUH1480)
N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstru
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
A:Accession: JC5620
R:Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.
Biochem. Biophys. Res. Commun. 236, 44-49, 1997
A:Title: The complete coding sequence of hepatitis C virus genotype 5a, the predomina
A:Reference number: JC5620; MUID:97366593; PMID:9223423

RESULT 9
JC5620
genome polyprotein - hepatitis C virus (isolate EUH1480)
N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstru
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
A:Accession: JC5620
R:Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.
Biochem. Biophys. Res. Commun. 236, 44-49, 1997
A:Title: The complete coding sequence of hepatitis C virus genotype 5a, the predomina
A:Reference number: JC5620; MUID:97366593; PMID:9223423

genome polyprotein - hepatitis C virus (isolate EUH1480)
N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstru
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
A:Accession: JC5620
R:Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.
Biochem. Biophys. Res. Commun. 236, 44-49, 1997
A:Title: The complete coding sequence of hepatitis C virus genotype 5a, the predomina
A:Reference number: JC5620; MUID:97366593; PMID:9223423

```
Query Match      65.1%  Score 672;  DB 1;  Length 3033;
Best Local Similarity 68.7%  Pred. No. 8.4e-54;
Matches 123;  Conservative 24;  Mismatches 32;  Indels 0;  Gaps 0;
```

```

QY      19  TAYAAQTGEECCQETSGTRKKNQVEVQIVSTAAQTFLATCINGVCWTVYHGAGTRT  78
Db      1034  TAYTQOTRGELLGAIIVSLTGRDKNQAGQVQLSSVTQTFLGTSISGVLWTVYHGAGNK  1093

QY      79  IASPGKPVQIMYTNVDKDLVGHAPQGSRSILTPCTCGSSDLKLVTRHADVIPIVRRGDGR  138
Db      1094  IASPGKPVQIMYTSAGDVLGVWSPSPGKSLDPCCTCGADVLYLVTRNADVIPIVRRKDDRR  1153

QY      139  GSLLSPRPISYLKSGSGGPLLCPAGHAGVIFPRAAVCTRGVAKAVDPIPVESLETTMBSP  197
Db      1154  GALLSPRLSTLTKSGSGGPVLCRGHAVGLFRAAVCARGVAKSIDPIPVESLVDATRPT  1212

RESULT 12
T08841
polyprotein - douroucouli hepatitis GB virus A
C:Species: douroucouli hepatitis GB virus A
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000
C:Accession: T08841
R:Erker, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.
J. Gen. Virol. 79, 41-45, 1998
A:Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.
A:Reference number: Z16486; MUID:98120818; PMID:9460920
A:Accession: T08841
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-3005 <ER>
A:Cross-references: EMBL:AF023425; NID:g2828599; PIDN:AAC40502.1; PID:g2828600
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: polyprotein

```

```

Qy 53 TAAQTFLATCVCVWTVTHGACTRTIAGKGPVIOMYTNDVKDLGVWPAQGSRLTPC 112
Db 1006 TSTRSGKTCVNGVWYTTTHGSNARTLAAQMGPVNSRWMSASDDVAVYPLPVCAKCLEPC 1065
Qy 113 TCGSSDLYLWTRHADYIPVRRRGDSRGSLLS-----PRPISYLGSGGGPLLCP 161
Db 1066 KCOQGVWVI-----RND--GALCHGTGLRTVELDPAELCDPRGSGSFLICD 1112
Qy 162 AGHAGVIFRAAVCTRG-----VAKAVDFIPVESLETTMRSP 197
Db 1113 EGHAVGML-ISVLRGRSVTGIRYTKPWETLPREAITHTEAPP 1154

RESULT 14
B71360
hypothetical protein TP0136 - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C:Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 05-Nov-1999
C:Accession: B71360
R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.;
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; M
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-386, 1998
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A:Reference number: A71250; MUID:98332770; PMID:9665876
A:Accession: B71360
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-495 <COL>
A:Cross-references: GB:AE0001199; GB:AE000520; NID:g3322402; PIDN:AAC65137.1; PID:g332
A:Experimental source: strain Nichols
C:Genetics:
A:Gene: TP0136

```

[illegible]

A:Cross-references: GDB:4573993
A:Map position: l4q32.1-l4q32.1
C:Superfamily: unassigned WD repeat proteins; WD repeat homology
F:148-181/Domain: WD repeat homology <WD1>
F:414-447/Domain: WD repeat homology <WD2>

Query Match 7.9%; Score 82; DB 2; Length 452;
Best Local Similarity 25.3%; Pred. No. 7.1;
Matches 42; Conservative 13; Mismatches 47; Indels 64; Gaps 9;

Qy	68	WTVVHGACTRTIASPKGPIOMYTINVDKLVGWPAPOGSRSL-----TPCTCGSSDLYLV	122
Db	197	WNEWH-----PRAPVLLAGT-ADGNTWMMKVPNQDKCTFOGPNCPATCGR-----	240
Qy	123	TRHADVIPVRRR---GDSRGS-----LLSPRPISYLGSSG--GPLLCPA-----	162
Db	241	-----VLPDGKRAVVGYEDGTIRIWLKQGSPIHVLKGTGEGHGLTCVAAANQDGLILT	295
Qy	163	-----GHAVGIFR-----AAVCTRGVAKAVDFIPVESL	190
Db	296	GSVDCQAKLVSATGKVVGVFRPETVASQPSLGECESESNSVESL	341

Search completed: August 30, 2003, 19:20:27
Job time : 17.2134 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 18:01:52 ; Search time 9.75674 Seconds
(without alignments)
949.524 Million cell updates/sec

Title: US-09-965-594-14

Perfect score: 1032

Sequence: 1 MKKKGSVIVGRINLSGDTA.....VAKAVDFIPVESLETTMRSP 197

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	901.5	87.4	3011	1 POLG_HCV1	P26664 h genome po
2	886.5	85.9	3011	1 POLG_HCVH	P27958 h genome po
3	854.5	82.8	3010	1 POLG_HCVTW	P29846 h genome po
4	850.5	82.4	3010	1 POLG_HCVBK	P26663 h genome po
5	850.5	82.4	3010	1 POLG_HCVJA	P26662 h genome po
6	844.5	81.8	3010	1 POLG_HCVJT	Q00269 h genome po
7	674	65.3	3033	1 POLG_HCVJ6	P26660 h genome po
8	672	65.1	3033	1 POLG_HCVJ8	P26661 h genome po
9	85.5	8.3	485	1 Y136_TREPA	O83172 treponema p
10	85	8.2	321	1 HHOA_ARATH	Q9sel7 arabidopsis
11	82	7.9	452	1 AAMP_HUMAN	Q13685 homo sapien
12	80.5	7.8	437	1 DEGL_ARATH	O22609 arabidopsis
13	78.5	7.6	209	1 PAAD_PSEAE	Q9hx08 pseudomonas
14	77.5	7.5	2663	1 CENE_HUMAN	Q02224 homo sapien
15	76	7.4	911	1 TBIL_NEIMB	Q09056 neisseria m
16	75.5	7.3	786	1 SNIL_HUMAN	P57059 homo sapien
17	75	7.3	603	1 ENV_RSVP	P70289 rous sarcom
18	75	7.3	1705	1 TPPO_MOUSE	P70289 mus musculu
19	74.5	7.2	415	1 ZP3_RABIT	P40833 oryctolagus
20	74.5	7.2	776	1 HYPF_AZOVI	P40596 azotobacter
21	74	7.2	1165	1 POLG_GALV	P21414 gibbon ape
22	73.5	7.1	263	1 GRAB_MOUSE	O35205 mus musculu
23	73.5	7.1	661	1 INV8_DAUCA	P80065 daucus caro
24	73	7.1	253	1 CAC3_BOVIN	P05805 bos taurus
25	73	7.1	259	1 IBP1_HUMAN	P08833 homo sapien
26	72.5	7.0	257	1 GRAM_HUMAN	P51124 homo sapien
27	72.5	7.0	706	1 TRFE_HORSE	P27425 equus cabal
28	72	7.0	659	1 VST2_HEVNE	Q03500 hepatitis e
29	72	7.0	1527	1 CALH_MOUSE	P39061 mus musculu
30	71.5	6.9	248	1 GRAD_MOUSE	P11033 mus musculu
31	71.5	6.9	248	1 TRY1_CHICK	Q90627 gallus gall
32	71.5	6.9	408	1 SEPR_THESK	P80146 thermus sp.
33	71	6.9	336	1 UL16_EBV	P03221 epstein-bar

RESULT 1

ID	POLG_HCV1	STANDARD	PRT	3011 AA
AC	P26664			
DT	01-AUG-1992 (Rel. 23, Created)			
DT	01-AUG-1992 (Rel. 23, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].			
OS	Hepatitis C virus (isolate 1) (HCV).			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.			
OX	NCBI_TaxID=11104;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=91172826; PubMed=1848704;			
RA	Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C., Gallegos C., Colt D., Medina-Selby A., Barr P.J., Weiner A.J., Bradley D.W., Kuo G., Houghton M.			
RA	*Genetic organization and diversity of the hepatitis C virus.*;			
RT	Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).			
RL	-!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.			
CC	NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.			
CC	-!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.			
CC	-!- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate + (RNA)(N).			
CC	-!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND RNA.			
CC	-!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@lsb-sib.ch).			
CC	EMBL; M62321; AAA45676.1; -			
DR	PDB; A39166; GNWVC3.			
DR	PDB; 1ALV; 16-FEB-99.			
DR	PDB; 1HEI; 25-NOV-98.			
DR	MEROPS; S29.001; -			
DR	MEROPS; U39.001; -			
DR	InterPro; IPR001410; DEAD.			
DR	InterPro; IPR002522; HCV_capsid.			

P54748 rattus norv
Q01996 neisseria g
P18614 rattus norv
Q09550 caenorhabdl
P40313 homo sapien
P08531 aquifex aeo
P29359 feline leuk
P33426 hepatitis e
P19532 hepatitis e
Q987n3 rhizobium l
P20575 pseudomonas

ALIGNMENTS

34 71 6.9 844 1 CN4A_RAT
35 71 6.9 915 1 TBPI_NEIGO
36 71 6.9 1180 1 ITAL_RAT
37 71 6.9 1240 1 YQJ3_CAEEL
38 70.5 6.8 264 1 CTRL_HUMAN
39 70.5 6.8 443 1 FLIT_AQUAE
40 70.5 6.8 642 1 ENV_FLVGL
41 70.5 6.8 660 1 VST2_HEVBU
42 70.5 6.8 660 1 VST2_HEVPA
43 70.5 6.8 743 1 TFE3_HUMAN
44 70 6.8 326 1 PANE_RHILO
45 70 6.8 349 1 TRPD_PSEPU

DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRP.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; Helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRP; 1.
 DR ProDom: PD186062; DEXdc; 1.
 DR SMART: SM00487; DEXdc; 1.
 DR PolyProtein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3011
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 476 476
 FT CARBOHYD 532 532
 FT CARBOHYD 540 540
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2364 2364
 FT CARBOHYD 2789 2789
 SQ SEQUENCE 3011 AA; 65P8C9447FCESAF9 CRC64;

Query Match 87.48; Score 901.5; DB 1; Length 3011;
 Best Local Similarity 85.84; Pred. No. 5.5e-77;
 Matches 175; Conservative 9; Mismatches 11; Indels 9; Gaps 1;
 3 KKGSVIVGRIN-----LSGDTAYAAQOTRCEGCQETSQTGRDKNQVEGEVQIVST 53

DB 1005 RRGREILLGPADGMVSKGWRLLAPITAYAAQOTRGLGCIITSLTGRDKNQVEGEVQIVST 1064
 QY 54 AAQTFLATCINGVCWTYVHGAGTETIASPKGPVIQMTYNDKDLVGPAPQGSRLTPTCT 113
 DB 1065 AAQTFLATCINGVCWTYVHGAGTETIASPKGPVIQMTYNDKDLVGPAPQGSRLTPTCT 1124
 QY 114 CGSSDLVLTTHADVIPVRRGDSRGLSPRPISYLKSGSGGGLLCPAGHAGVIFRAAV 173
 DB 1125 CGSSDLVLTTHADVIPVRRGDSRGLSPRPISYLKSGSGGGLLCPAGHAGVIFRAAV 1184
 QY 174 CTRGVAKAVDFIPVESLETTMRSP 197
 DB 1185 CTRGVAKAVDFIPVENLETTMRSP 1208
 RESULT 2
 POLG_HCVH STANDARD; PRT; 3011 AA.
 AC P27958;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate H) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID-111108;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92052256; PubMed-1658800;
 RA Inchauspe G., Zebede S., Lee D.H.H., Sugitani M., Nasoff M.,
 RA Prince A.M.;
 RT "Genomic structure of the human prototype strain H of hepatitis C
 virus: comparison with American and Japanese isolates.";
 Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
 RL [2]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
 RX MEDLINE-97331322; PubMed-9187654;
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
 RT "Structure of the hepatitis C virus RNA helicase domain.";
 Nat. Struct. Biol. 4:463-467(1997).
 RL [3]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
 RX MEDLINE-98154321; PubMed-9493270;
 RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
 RA Murcko M.A., Lin C., Caron P.R.;
 RT "Hepatitis C virus NS3 RNA helicase domain with a bound
 oligonucleotide: the crystal structure provides insights into the mode
 of unwinding.";
 Structure 6:89-100(1998).
 RL [4]
 CC -1- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
 CC -1- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
 CC -1- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
 CC ACTIVATION OF NS3.
 CC -1- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
 CC -1- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC {RNA}(N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPID-PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.

Matches 171; Conservative 10; Mismatches 14; Indels 9; Gaps 1;

QY 3 KKSQVIVGRIN-----LSGDTAYAOOTREGCQTSOTGRDKNKQVEGEVIVST 53
D 1005 RRGQIILGPADGMVSKGWRLLAPITAYAOOTREGLLCIIITSLTGRDKNKQVEGEVIVST 1064
QY 54 AAOITFLATCINGVWTVYHCAGTTRTIASPKGSPVQMYTNDKLVGMPAPQGSRSITPCT 113
D 1065 ATOITFLATCINGVWTVYHCAGTTRTIASPKGSPVQMYTNDKLVGMPAPQGSRSITPCT 1124
QY 114 CGSDDLVLVTRHADVIPVRRGDSRGLSPRISYILKSGSGGLLCPAGHAYGIFRAAV 173
D 1125 CGSDDLVLVTRHADVIPVRRGDSRGLSPRISYILKSGSGGLLCPCTCHAVGLPRAAV 1184
QY 174 CTRGVAKAVDFIPVSELETTMRSP 197
D 1185 CTRGVAKAVDFIPVENLETTMRSP 1208

RESULT 3
POLG_HCVTW
ID POLG_HCVTW STANDARD; PRT: 3010 AA.
AC P29846;
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate Taiwan) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31645;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230206; PubMed=1314449;
RA Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;
FT "The Taiwanese hepatitis C virus genome: sequence determination and
FT mapping the 5' terminus of viral genomic and antigenomic RNA.";
RL Virology 188:102-113(1992).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
CC {RNA}(N).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M84754; ; NOT ANNOTATED_CDS.
DR PIR: A40244; GNAVTV.
DR PDB: 1N64; 25-FEB-03.
DR PDB: 1NS3; 08-APR-98.
DR MEROPS: S29.001; .
DR MEROPS: U39.001; .
DR InterPro: IPR001410; DEAD.

DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RDRP.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RDRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
FT CHAIN 116 191 CORE PROTEIN (POTENTIAL).
FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN 1007 1615 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT CHAIN 2014 3010 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT CHAIN 347 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT TRANSMEM 347 369 POTENTIAL.
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDFE215 CRC64;

Query Match 82.8%; Score 854.5; DB 1; Length 3010;
Best Local Similarity 78.9%; Pred. No. 1.6e-72;
Matches 161; Conservative 18; Mismatches 16; Indels 9; Gaps 1;

DR InterPro: IPR002166; HCV_RdRp.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NSI; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRp; 1.
 DR ProDom: PD186062; HCV_NSI; 1.
 DR SMART: SM00487; DEXDC; 1.
 KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Coat protein; Hydrolyase; Serine protease.
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 532 532
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2788 2788
 SQ SEQUENCE 3010 AA; 327017 MW; AA993794F46DB185 CRC64;

Query Match 82.4%; Score 850.5; DB 1; Length 3010;
 Best Local Similarity 76.5%; Pred. No. 3.8e-72;
 Matches 156; Conservative 23; Mismatches 16; Indels 9; Gaps 1;

QY 3 KGSWVIVGRINLSGD-----TAAQOTRGECCQTSOTGRDKNOVEGEVOIVST 53
 DB 1005 RGKEILLGADSGEGCGWRLLAPITASQOTRGLGCIITSLTGRDKNQVDGEVQLST 1064
 QY 54 AAQFLATFCINGVCWTVYHGAGTFTIAPKPGVITQMTYNDKDLVGPAPQGSRLTPTCT 113
 DB 1065 ATQSFATCVNGVCWTVYHGAGTFTIAPKPGVITQMTYNDKDLVGPAPQGSRLTPTCT 1124
 QY 114 CGSSDLYLVTRHADYVIVRRRGRSGSLSPRPISYLKGSGGGLLCPAGHANGVIFRAAV 173
 DB 1125 CGSSDLYLVTRHADYVIVRRRGRSGSLSPRPISYLKGSGGGLLCPAGHANGVIFRAAV 1184

QY 174 CTRGVAKAVDFIPVESLETTMRSP 197
 DB 1185 CTRGVAKAVDFIPVESMETTMRSP 1208

RESULT 6

POLG_HCVJT STANDARD; PRT; 3010 AA.
 AC Q00269;
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-APR-1993 (Rel. 25, Last annotation update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-JT) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=31642;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92295714; PubMed=1318627;
 RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,
 RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
 RT "Molecular cloning of hepatitis C virus genome from a single Japanese
 RT carrier: sequence variation within the same individual and among
 RT infected individuals";
 RL Virus Res. 23:39-53(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC {RNA}(N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D11168; BAA01943.1; .
 DR PIR; A45573; A45573
 DR PDB; 1AIQ; 25-MAR-98.
 DR PDB; 1JXP; 14-JAN-98.
 DR MEROPS; S29.001; .
 DR MEROPS; U39.001; .
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV capsid.
 DR InterPro: IPR002521; HCV core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NSI.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRp.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.


```
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXdc; 1.
KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural removed from capsid protein C BY THE
FT INIT_MET 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 733
FT CHAIN 734 1010
FT CHAIN 1011 1619
FT CHAIN 1620 1866
FT CHAIN 1867 2017
FT CHAIN 2018 3033
FT TRANSMEM 347 369
FT ACT_SITE 1087 1087
FT ACT_SITE 1111 1111
FT ACT_SITE 1169 1169
FT NP_BIND 1234 1241
FT SITE 1320 1323
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 233 233
FT CARBOHYD 299 299
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 477 477
FT CARBOHYD 534 534
FT CARBOHYD 542 542
FT CARBOHYD 558 558
FT CARBOHYD 578 578
FT CARBOHYD 627 627
FT CARBOHYD 649 649
FT CARBOHYD 1091 1091
FT CARBOHYD 2038 2038
FT CARBOHYD 2359 2359
FT CARBOHYD 2811 2811
FT CARBOHYD 3033 3033
FT SEQUENCE 3033 AA; 330177 MW; 1A173E7E3381FD1A CRC64;

Query Match 65.1%; Score 672; DB 1; Length 3033;
Best Local Similarity 68.7%; Pred. No. 3.1e-55;
Matches 123; Conservative 24; Mismatches 32; Indels 0; Gaps 0;

QY 19 TAYAAQTRGEGCGQETSGTGRDKNQVEGEVOIVSTAQTFLATCINGVCTVYHGAGTRT 78
DB 1034 TAYTQOTRGLLGAIVVSLTGRDKNQVEGEVOIVSTAQTFLATCINGVCTVYHGAGNKT 1093
QY 79 IASPKGPVIOYTNVDKDLVCPAQGSRSUTPCTCGSSDLVLYTRHADVIVRRRGDSR 138
DB 1094 LAGPKGPVIOYTNVDKDLVCPAQGSRSUTPCTCGSSDLVLYTRHADVIVRRRGDSR 1153
QY 139 GSLSPRISYKSGSGPLCPAGHVGIPRAAVCTRGVAKAVDFIPVSELTMRSP 197
DB 1154 GALLSPRLSLTKGSGGVLCSRGHVGIPRAAVCTRGVAKAVDFIPVSELTMRSP 1212

RESULT 9
Y136_TREPA
ID Y136_TREPA STANDARD; PRT; 485 AA.
AC 083172;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical lipoprotein TP0136 precursor.
GN TP0136.
OS Treponema pallidum.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;

[1]
RN SEQUENCE FROM N.A.
RP STRAIN=Nichols;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khatak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete.
RL Science 281:375-388(1998).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC (potential).
CC -!- SIMILARITY: BELONGS TO THE TP013X FAMILY OF LIPOPROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF001199; AAC65137.1; ALT_INIT.
CC TIGR: TP0136; -.
CC Hypothetical protein; Lipoprotein; Membrane; Signal;
CC Complete proteome.
CC SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 485 HYPOTHETICAL LIPOPROTEIN TP0136.
FT LIPID 24 24 N-ACYL DIGLYCERIDE (POTENTIAL).
FT DOMAIN 164 178 GLY/SER-RICH.
FT DOMAIN 196 210 GLY/SER-RICH.
FT DOMAIN 253 267 GLY/SER-RICH.
FT DOMAIN 318 327 POLY-SER.
FT DOMAIN 444 447 POLY-SER.
FT SEQUENCE 485 AA; 48984 MW; C7A4CEDC7DC5CED CRC64;

Query Match 8.3%; Score 85.5; DB 1; Length 485;
Best Local Similarity 23.4%; Pred. No. 1.3;
Matches 50; Conservative 17; Mismatches 76; Indels 71; Gaps 10;

QY 16 SGTAYV-----QOTRGECCQETSGTGRDKNQVEGEVOIVSTAQTFLATCI- 63
DB 54 AGSKLYATNGRWELNLTGSGWQVSSSVPTDSK-----KVSATDGTFTVLACVP 108
QY 64 -NGCVTVYHGAG---TRTIASPKGPVIOYTNVDKDLVG-----WPAPOGSRSLTPCT 113
DB 109 GTGVYKHCYVNGAGSSSTGTGTASPTETCSQAT-----LVGGTSKPFVLVPGGTGNGMCG 164
QY 114 C-----GSSDLYLVTRHADVIV-----VRRRGDSRGLSLSPRISYK----- 151
DB 165 CGGGGGGSSSSSSCHIHVLVPGGTGNGMCGCGGGGGSSSSSSCHIKVNTDEOFL 224
QY 152 -----GSSGGPLLCFAGHVG 167
DB 225 DMGEGYVVTTKHLTKNGSSSAGPACQCGGGGG 258

RESULT 10
HHOA_ARATH
ID HHOA_ARATH STANDARD; PRT; 321 AA.
AC 09SEL7; 049507;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Protease HhoA, chloroplast precursor (EC 3.4.21.-).
GN HHOA OR AT4G18370 OR F28J12.30.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
```


J. Biol. Chem. 273:7094-7098(1998).

[2]

SEQUENCE FROM N.A.
STRAIN=cv, Columbia;
MEDLINE=2036309; PubMed=10907853;
Kaneke T., Katoch T., Sato S., Nakamura A., Asamizu E., Tabata S.;
*Structural analysis of Arabidopsis thaliana chromosome 3. II,
Sequence features of the 4,251,695 bp regions covered by 90 Pl. TAC
and BAC clones.*;
DNA Res. 7:217-221(2000).
[3]
SEQUENCE OF 104-118:
STRAIN=cv, Columbia;
Kieselbach T., Bystedt M., Schroeder W.P.:
Submitted (JUL-2000) to the SWISS-PROT data bank.
-|- FUNCTION: SERINE PROTEASE THAT IS REQUIRED AT HIGH TEMPERATURE.
MAY BE INVOLVED IN THE DEGRADATION OF DAMAGED PROTEINS. IN VIVO,
CAN DEGRADE BETA-CASEIN.
-|- ENZYME REGULATION: INHIBITED BY PHENYLMETHYLSULFONYL FLUORIDE AND
O-PHENANTHROLINE.
-|- SUBCELLULAR LOCATION: BOUND TO LUMINAL SIDE OF THE THYLAKOID
MEMBRANE.
-|- INDUCTION: By heat shock.
-|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
-|- SIMILARITY: Contains 1 PDZ/DHR domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

EMBL: AF028842; AAC39436.1; .
EMBL: AP000371; BAB02539.1; .
EMBL: AP001302; BAB02539.1; JOINED.
MEROPS: S01.279; .
InterPro: IPR001478; PDZ.
InterPro: IPR001940; Protease2C.
InterPro: IPR001254; Ser_protease_Try.
Pfam: PF00595; PDZ; 1.
Pfam: PF00089; trypsin; 1.
PRINTS: PR00834; PROTEASES2C.
SMART: SM00228; PDZ; 1.
PROSITE: PS0106; PDZ; 1.
Hydrolase: Serine protease;
TRANSIT ?
FT TRANSIT 7 103
FT CHAIN 104 437
FT DOMAIN 152 321
FT DOMAIN 324 421
FT ACT_SITE 171 171
FT ACT_SITE 201 201
FT ACT_SITE 280 280
FT CONFLICT 12 23
FT CONFLICT 36 36
FT CONFLICT 54 54
FT CONFLICT 60 60
FT CONFLICT 64 64
FT CONFLICT 68 69
FT CONFLICT 355 355
FT CONFLICT 381 381
FT CONFLICT 416 416
FT SEQUENCE 437 AA; 46212 MW; 1497B1AB3F5FF2A4 CRC64;

Query Match 7.8%; Score 80.5; DB 1; Length 437;
Best Local Similarity 25.68; Pred.No. 3;
Matches 44; Conservative 18; Mismatches 55; Indels 55; Gaps 7;

QY 70 YVHGAGTTIASPKGVPIQM; TNVDKLVGWM-----PA 102

```

Db 150 VPOGSGGFWDKQGHVHTNYHVRGASDLRVTLADQTTEDAKVVGFDQDKDVAVLRLDA 209
QY 103 PQGSRRLTPTCTGSSDLXLY-----TRHADVTPVRRRGDSRGSLSPRP 147
Db 210 PK--NKLRPVGVSAADLVGVKQVFAIGNPFLDHTLTITGVISGLRREIS--SAATGRPI 265
QY 148 SYL-----KGSGGGLLCPAGHAVGIFRAAVCTRGVAKAVDF-IPVESL 190
Db 266 QDVQTDAAINFGNSGGPLDSSGTLIGINTAIYSPGASSGVGFSIPVDTV 317

RESULT 13
PAAD_PSEAE
ID PAAD_PSEAE STANDARD; PRT; 209 AA.
AC Q9HX08;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable aromatic acid decarboxylase (EC 4.1.1.-).
GN PA4019.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Hickey C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -1- SIMILARITY: BELONGS TO THE POLYPRENYL P-HYDROXYBENZOATE /
CC PHENYLACRYLIC ACID DECARBOXYLASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF004818; AAC07406.1; -
CC PIR; H83144;
CC InterPro; IPR003382; Flavoprotein.
CC Pfam; PF02441; Flavoprotein; 1.
CC KW Hypothetical protein; Lyase; Decarboxylase; Complete proteome.
CC SEQUENCE 209 AA; 22367 MW; 01FD081CC495D3F6 CRC64;

Query Match 7.68; Score 78.5; DB 1; Length 209;
Best Local Similarity 26.5%; Pred. No. 2.3;
Matches 50; Conservative 16; Mismatches 56; Indels 67; Gaps 11;

QY 43 QVEGEVO-IVSTAOTFLATCINGVCVTYVHGACTRTTIASPKG----- 85
Db 29 QEEREVFLISKAAQLVWAT-----ETDVALPAKPAQMAFLTEYCGAAG 74
QY 86 VIQMTYNDKDLVGWPAQGRSLTP-----CTGSSDL-----YLVTHADVIPV 131
Db 75 QIRVEGQND-----WMAPPASGSAPNAVMICPGSTGTLTSAVATGACNNLIERAADVALK 129
QY 132 RRGDSRGSLSPR--PIS-----YLGSGGGLLCPAGHAVGIFRAAVCTRGVAKAVD 193
Db 130 ER-----RPLVLPREAPFSSITHLENMLKLSNLGAVILPA--AFGFYHQ-----POSVEDLVD 180
QY 184 FIPVESLET 192

```

```

Db 181 FVVARILNT 189

RESULT 14
CENE_HUMAN
ID CENE_HUMAN STANDARD; PRT; 2663 AA.
AC Q02224;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Centromeric protein E (CENP-E protein).
GN CENPE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93024922; PubMed=1406971;
RA Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;
RT "CENP-E is a putative kinetochore motor that accumulates just before
RT mitosis.";
RL Nature 359:536-539(1992).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=95196755; PubMed=7889940;
RA Thrower D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;
RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
RL microtubule motor.";
RL EMBO J. 14:918-926(1995).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=98437347; PubMed=9763420;
RA Chan G.K.T., Schaar B.T., Yen T.J.;
RT "Characterization of the kinetochore binding domain of CENP-E reveals
RT interactions with the kinetochore proteins CENP-F and hBUBR1.";
RL J. Cell Biol. 143:49-63(1998).
CC -1- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
CC KINETOCORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
CC AND/OR SPINDLE ELONGATION.
CC -1- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.
CC -1- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCORES DURING
CC CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z15005; CAAT8727.1; -
CC PIR; S28261; S28261.
CC HSP; P17119; 3KAR.
CC Genew; HGNC:1856; CENPE.
CC GK; Q02224; -
CC MIM; 117143; -
CC GO; GO:0005699; C:kinetochore; TAS.
CC GO; GO:0005634; C:nucleus; TAS.
CC GO; GO:0008350; F:kinetochore motor activity; TAS.
CC GO; GO:0000067; F:DNA replication and chromosome cycle; TAS.
CC GO; GO:0007079; P:mitotic chromosome movement; TAS.
CC GO; GO:0007080; P:mitotic metaphase plate congression; TAS.
CC InterPro; IPR001752; kinesin_motor.
CC Pfam; PF00225; kinesin; 1.
CC PRINTS; PR00380; KINESINHEAVY.
CC SMART; SM00129; KISC; 1.
CC PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
CC PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.

```

KW Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;

KW Cell cycle; Centromere.

FT DOMAIN 1 335 KINESIN-MOTOR.

FT DOMAIN 336 2471 COILED COIL (POTENTIAL).

FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).

FT NP_BIND 86 93 ATP (BY SIMILARITY).

SQ SEQUENCE 2663 AA; 312087 MW; CEFCL3880C8C8CB8 CRC64;

Query Match 7.5%; Score 77.5; DB 1; Length 2663;

Best Local Similarity 24.0%; Pred. No. 56;

Matches 41; Conservative 15; Mismatches 52; Indels 63; Gaps 8;

QY 32 QETSGTRDKNQVEGEQIVSTAQT-----LATCINGVCWTYHAGTRTIA 80

Db 2523 CQNEQLIKQKNLLSNQHLNNEVTKWTKRLKRAHKQVTCE----- 2566

QY 81 SPKGPVIOYTNVDKLVGHPAPOGSRSLTPCTCGSSDLVLTTRHADVIPVRRG-----D 136

Db 2567 SPKSPKVTGTASKKK-----OITPSQCKERNL-----QDPVPKSPKSCFFD 2608

QY 137 SRG-SLLSPRISYLKSSGGPLCPAGHAGVIFRAAVCTRGVAKAVDFIP 186

Db 2609 SRKSLSPSPHVPVFDNSSLG--LCPEVQNAG-----AESVDSQP 2646

RESULT 15

TB1L_NEIMB

ID TB1L_NEIMB STANDARD; PRT; 911 AA.

AC Q09056;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Transferrin-binding protein 1 precursor.

GN TBPL.

OS Neisseria meningitidis (serogroup B).

OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;

OC Neisseriaceae; Neisseria.

OX NCBI_TaxID=491;

RI [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC STRAIN=CCUG 37608 / M982 / Serogroup B / Serotype 9;

RX MEDLINE=93345825; PubMed=8344530;

RA Legrain M., Mazarin V., Irwin S.W., Bouchon B., Quentin-Willet M.-J.,

RA Jacobs E., Schryvers A.B.;

RT *Cloning and characterization of Neisseria meningitidis genes

RT encoding the transferrin-binding proteins Tbp1 and Tbp2. #;

RL Gene 130/73-80(1993).

CC -!- FUNCTION: ACTS AS A TRANSFERRIN RECEPTOR AND IS REQUIRED FOR

CC TRANSFERRIN UTILIZATION.

CC -!- SUBCELLULAR LOCATION: Outer membrane.

CC -!- INDUCTION: By iron starvation.

CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

CC -----

CC EMBL; Z15130; CAA78833.1; .

DR PIR; JN0821; JN0821.

DR InterPro; IPR000531; TonB_boxC.

DR Pfam; PF00593; TonB_dep_Rec1.

DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.

DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.

KW Outer membrane; Receptor; Signal; TonB box.

FT SIGNAL 1 24

FT CHAIN 25 911 TRANSFERRIN-BINDING PROTEIN 1.

FT SITE 38 45 TONB_BOX.

FT SITE 894 911 TONB_C-TERMINAL_BOX.

SQ SEQUENCE 911 AA; 101631 MW; 99283ABAE0B773E5 CRC64;

Query Match 7.4%; Score 76; DB 1; Length 911;

Best Local Similarity 26.4%; Pred. No. 22;

Matches 60; Conservative 20; Mismatches 81; Indels 66; Gaps 13;

QY 2 KKGSVIVIGRINLSGDTAYAAQ-----TRGEGCQETSQ-----TGRDKNQ- 43

Db 50 RRDNEVTGLGLVKVTADTLSKEQVLDIRDLTRYDPGLAVVEQGRGASSGYSIRGMDKNRV 109

QY 44 ---VEGEVQIVSTAQTFLATCINGVCWTYHAGTRTIA SPKGPVIOY-TYNDKDLVG 99

Db 110 SLTVDGLAQIOSYTAQAAL-----GGTET-AGSSGAINIEIENYKAVEIS 154

QY 100 WPAQGSRSRLTPCTCGSSDL-----YLVTRHADVIPVRRR-----GDSRGSLSP 144

Db 155 ----KGSNSVEQ---GSGALAGSVAFTKTADDVIGEGRWGIOSKATAYSGKNRGLTQS- 206

QY 145 RPISVLKSSGG--PLLCPAGHAGVIFRA-AVCTRGVAKAVDFIPVE 188

Db 207 ---TALAGRIGGAELLIHTGRRAGEIRAHEDAGRGVQSFNRLVPVE 250

Search completed: August 30, 2003, 19:13:45

Job time : 10.7567 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 19:00:22 : Search time 37.5921 Seconds
(without alignments)
1352.314 Million cell updates/sec

Title: US-09-965-594-14

Perfect score: 1032

Sequence: 1 MKKKGSVVIGRINLSGDTA.....VAKAVDFIPVESLETTMRSP 197

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23.*

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mhc.*

8: sp_organelle.*

9: sp_phase.*

10: sp_plant.*

11: sp_todent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

15: sp_rvirus.*

16: sp_bacteriap.*

17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	910.5	88.2	4040	12 Q91FH8	Q91FH8 mucosal dis
2	901.5	87.4	2436	12 Q81756	Q81756 hepatitis c
3	901.5	87.4	3011	12 Q91FE5	Q91FE5 hepatitis c
4	900.5	87.3	3011	12 Q03453	Q03453 hepatitis c
5	896.5	86.9	3011	12 Q36579	Q36579 hepatitis c
6	896	86.8	181	12 Q91RR8	Q91RR8 hepatitis c
7	896	86.8	181	12 Q91RT5	Q91RT5 hepatitis c
8	894	86.6	181	12 Q91RR5	Q91RR5 hepatitis c
9	893	86.5	181	12 Q81RE3	Q81RE3 hepatitis c
10	893	86.5	181	12 Q91RS1	Q91RS1 hepatitis c
11	893	86.5	181	12 Q91R08	Q91R08 hepatitis c
12	893	86.5	181	12 Q91RT1	Q91RT1 hepatitis c
13	892.5	86.5	3011	12 Q91LS8	Q91LS8 hepatitis c
14	891	86.3	181	12 Q91RR6	Q91RR6 hepatitis c
15	891	86.3	181	12 Q91RS9	Q91RS9 hepatitis c
16	890	86.2	181	12 Q91RR2	Q91RR2 hepatitis c

17	890	86.2	181	12 Q91RS3	Q91RS3 hepatitis c
18	889.5	86.2	3011	12 Q9DIT6	Q9DIT6 hepatitis c
19	889.5	86.2	3011	12 Q36608	Q36608 hepatitis c
20	889.5	86.2	3015	12 Q9PWX5	Q9PWX5 hepatitis c
21	889.5	86.2	3015	12 Q9PW09	Q9PW09 hepatitis c
22	889	86.1	181	12 Q91RT4	Q91RT4 hepatitis c
23	889	86.1	181	12 Q91RS8	Q91RS8 hepatitis c
24	889	86.1	181	12 Q91RT3	Q91RT3 hepatitis c
25	889	86.1	181	12 Q91RS5	Q91RS5 hepatitis c
26	889	86.1	181	12 Q91RS7	Q91RS7 hepatitis c
27	889	86.1	181	12 Q91RT0	Q91RT0 hepatitis c
28	887	85.9	181	12 Q91RS4	Q91RS4 hepatitis c
29	886	85.9	181	12 Q91RT6	Q91RT6 hepatitis c
30	885	85.8	181	12 Q91RT9	Q91RT9 hepatitis c
31	884	85.7	181	12 Q91RR4	Q91RR4 hepatitis c
32	884	85.7	181	12 Q91RR9	Q91RR9 hepatitis c
33	884	85.7	181	12 Q91RR0	Q91RR0 hepatitis c
34	883.5	85.6	3011	12 Q36609	Q36609 hepatitis c
35	882	85.5	181	12 Q91RR7	Q91RR7 hepatitis c
36	881	85.4	181	12 Q91RT2	Q91RT2 hepatitis c
37	881	85.4	181	12 Q91RR1	Q91RR1 hepatitis c
38	881	85.4	181	12 Q91RQ9	Q91RQ9 hepatitis c
39	881	85.4	181	12 Q91RS2	Q91RS2 hepatitis c
40	879	85.2	181	12 Q91RS6	Q91RS6 hepatitis c
41	878	85.1	181	12 Q91RT7	Q91RT7 hepatitis c
42	877	85.0	3011	12 Q36610	Q36610 hepatitis c
43	876	84.9	181	12 Q91RS0	Q91RS0 hepatitis c
44	876	84.9	181	12 Q91RT8	Q91RT8 hepatitis c
45	869.5	84.3	3010	12 Q9J3G9	Q9J3G9 hepatitis c

ALIGNMENTS

RESULT 1

ID	Q91FH8	PRELIMINARY;	PRT;	4040 AA.
AC	Q91FH8;			
DT	01-OCT-2000 (Tremblrel. 15, Created)			
DT	01-OCT-2000 (Tremblrel. 15, Last sequence update)			
DT	01-MAR-2003 (Tremblrel. 23, Last annotation update)			
DE	Genome polyprotein.			
OS	Mucosal disease virus.			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage: Flaviviridae;			
OX	Pestivirus.			
OC	NCBI_TaxID=11099;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	MEDLINE-20323484; PubMed-10864644;			
RA	Lai V.C., Zhong W., Skelton A., Ingravallo P., Vassilev V.,			
RA	Donis R.O., Hong Z., Lau J.Y.;			
RT	"Generation and characterization of a hepatitis C virus NS3 protease-			
RT	dependent bovine viral diarrhea virus.";			
RL	J Virol. 74:6339-6347(2000).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Lai V.C.H., Hong Z.;			
RL	Submitted (MAY/2000) to the EMBL/GenBank/DBJ databases.			
EMBL	AF268278; AAF82566.1; .			
HSSP	P26663; LXP.			
MEROPS	S31.001; .			
DR	InterPro: IPR000280; CDvir_endptsep80.			
DR	InterPro: IPR001410; DEAD.			
DR	InterPro: IPR004109; HCV_NS3.			
DR	InterPro: IPR002166; HCV_RdRP.			
DR	InterPro: IPR001650; Helicase_C.			
DR	InterPro: IPR001005; Myb_DNA_binding.			
DR	InterPro: IPR001568; RNase_T2.			
DR	InterPro: IPR007095; RNA_pol_DS_PS.			
DR	Pfam: PF02907; HCV_NS3; I.			
DR	Pfam: PF00271; helicase_C; 1.			
DR	Pfam: PF00998; Viral_RdRP; 1.			

```
DR PRINTS: PR00729; CDVENDOPTASE.
DR SMART: SM00487; DEXDC; 1.
DR SMART: SM00490; HELIC; 1.
DR PROSITE: PS00037; MYB_1; 1.
DR PROSITE: PS05007; RDRP_POSITIVE; 1.
DR PROSITE: PS05021; RDRP_VIRAL; 1.
DR PROSITE: PS00531; RNASE_T2_2; 1.
DR ATP-binding; Helicase; Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase.
SQ SEQUENCE 4040 AA: 453073 MW: ADE87791D055B9DC CRC64;

Query Match      88.2%  Score 910.5; DB 12; Length 4040;
Best Local Similarity 91.3%; Pred. No. 4.5e-83;
Matches 178; Conservative 5; Mismatches 9; Indels 3; Gaps 1;

QY 5 GSVVIVGRINLSD---TAYAQOTRGEEGCOETSGTGRDKNQVEGEVQIVSTAATFLAT 61
DB 10 GSVVIVGRIVLSGSGSITACQAQOTRGLLGCKITSLTGRDKNQVEGEVQIVSTAATFLAT 69
QY 62 CINGCVTVYHGAGTRTASPKGPVIQMTYNDKLVGWPAPQGSRLTPTCTCGSSDLYL 121
DB 70 CINGCVTVYHGAGTRTASPKGPVIQMTYNDKLVGWPAPQGSRLTPTCTCGSSDLYL 129
QY 122 VTRHADVIPVRRGDSRGSLLSPRPISYLKSGSGGPLLCPAGHAVGIFRAAVCTRGVAKA 181
DB 130 VTRHANVIPVRRGDSRGSLLSPRPISYLKSGSGGPLLCPAGHAVGIFRAAVCTRGVAKA 189
QY 182 VDFIPVESLETTMRS 196
DB 190 VDFIPVENLETTMRS 204

RESULT 2
Q81756 PRELIMINARY; PRT; 2436 AA.
AC Q81756;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21262212; PubMed=11369872;
RA Lanford R.E., Lee H., Chavez D., Guerra B., Brasky K.M.;
RT 'Infectious cDNA clone of the hepatitis C virus genotype 1 prototype
sequence.';
RL J. Gen. Virol. 82:1291-1297(2001).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL; AF271632; AAF81759.1; -.
DR HSP; P27958; 1A1V.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR000745; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS05007; RDRP_POSITIVE; 1.

DR PROSITE: PS05021; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
FT NON_TER 1 2436 2436
SQ SEQUENCE 2436 AA: 264734 MW: D7B9872900BE3125 CRC64;

Query Match      87.4%  Score 901.5; DB 12; Length 2436;
Best Local Similarity 85.8%; Pred. No. 1.9e-82;
Matches 175; Conservative 9; Mismatches 11; Indels 9; Gaps 1;

QY 3 KGSWIVGRIN-----LSGDTAYAQOTRGEEGCOETSGTGRDKNQVEGEVQIVST 53
DB 555 RRGREILLGPADGMVSGWELLAPITAYAQOTRGLLGCIITSLTGRDKNQVEGEVQIVST 614
QY 54 AAQTFLATCINGCVTVYHGAGTRTASPKGPVIQMTYNDKLVGWPAPQGSRLTPTCT 113
DB 615 AAQTFLATCINGCVTVYHGAGTRTASPKGPVIQMTYNDKLVGWPAPQGSRLTPTCT 674
QY 114 CGSSDLYLVTRHADVIPVRRGDSRGSLLSPRPISYLKSGSGGPLLCPAGHAVGIFRAAV 173
DB 675 CGSSDLYLVTRHADVIPVRRGDSRGSLLSPRPISYLKSGSGGPLLCPAGHAVGIFRAAV 734
QY 174 CTRGVAKAVDFIPVESLETTMRS 197
DB 735 CTRGVAKAVDFIPVENLETTMRS 758

RESULT 3
Q91FE5 PRELIMINARY; PRT; 3011 AA.
AC Q91FE5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21262212; PubMed=11369872;
RA Lanford R.E., Lee H., Chavez D., Guerra B., Brasky K.M.;
RT 'Infectious cDNA clone of the hepatitis C virus genotype 1 prototype
sequence.';
RL J. Gen. Virol. 82:1291-1297(2001).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL; AF271632; AAF81759.1; -.
DR HSP; P27958; 1A1V.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
```

DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRp; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR PROSITE: PS00190; CYTOCHROME_C; 1.
 DR PROSITE: PS50507; RDRP_POSITIVE; 1.
 DR PROSITE: PS50521; RDRP_VIRAL; 1.
 KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
 KW Hydrolase; Nonstructural protein; Polyprotein;
 KW RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3011 AA; 327124 MW; 2489CE74AC86AE58 CRC64;

Query Match 87.4%; Score 901.5; DB 12; Length 3011;
 Best Local Similarity 85.8%; Pred. No. 2.5e-82;
 Matches 175; Conservative 9; Mismatches 11; Indels 9; Gaps 1;

QY 3 KGSVVIVGRIN-----LSGDTAYAAQOTRGECCOETSOTGRDKNQVEGEVQIVST 53
 DB 1005 RRGREILLGPADGMVSKGWRLLAPITAYAAQOTRGLGCIITSLTGRDKNQVEGEVQIVST 1064

QY 54 AAQTFLATCINGVCWTVYHGAGTRTIAISPKGPVIQMTYNDKDLVGPAPQGSRSILTPCT 113
 DB 1065 AAQTFLATCINGVCWTVYHGAGTRTIAISPKGPVIQMTYNDKDLVGPAPQGSRSILTPCT 1124

QY 114 CGSSDLYLVTRHADVIPVRRGDSRGLSPRISYLYKSSGGPLCPAGHAVGIFRAAV 173
 DB 1125 CGSSDLYLVTRHADVIPVRRGDSRGLSPRISYLYKSSGGPLCPAGHAVGIFRAAV 1184

QY 174 CTRGVAKAVDFIPVESLETMRSP 197
 DB 1185 CTRGVAKAVDFIPVENLETTMRSP 1208

RESULT 4
 Q03463
 ID Q03463 PRELIMINARY; PRT; 3011 AA.
 AC Q03463;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Genome polyprotein.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxId=11103;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN-HC-J1;
 RX MEDLINE-91013116; PubMed-2170712;
 RA Okamoto H., Okada S., Sugiyama Y., Yotsumoto S., Tanaka T.,
 Yoshizawa H.;
 RT "The 5'-terminal sequence of the hepatitis C virus genome.";
 RL Jpn. J. Exp. Med. 60:167-177(1990).
 RN [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN-HC-J1;
 RX MEDLINE-92044440; PubMed-1658196;
 RA Okamoto H., Okada S., Sugiyama Y., Kurai K., Iizuka H., Machida A.,
 Miyakawa Y., Mayumi M.;
 RT "Nucleotide sequences of the genomic RNA of hepatitis C virus isolated
 from a human carrier: comparison with reported isolates for conserved
 and divergent regions.";
 RL J. Gen. Virol. 72:2697-2704(1991).
 RN [3]
 RN SEQUENCE FROM N.A.
 RC STRAIN-HC-J1;
 RX MEDLINE-93117120; PubMed-1335573;
 RA Okamoto H., Kanai N., Mishiro S.;

Full-length nucleotide sequence of a Japanese hepatitis C virus isolate (HC-J1) with high homology to USA isolates.;
 Nucleic Acids Res. 20:6410-6410(1992).
 [4]
 RN SEQUENCE FROM N.A.
 RC STRAIN-HC-J1;
 RA Okamoto H.;
 RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
 [5]
 RN SEQUENCE FROM N.A.
 RC STRAIN-HC-J1;
 RX MEDLINE-94174722; PubMed-7510436;
 RA Mink M., Benichou S., Madaule P., Tiollais P., Prince A.,
 Inchauspe G.;
 RT "Characterization and mapping of a B-cell immunogenic domain in hepatitis C virus E2 glycoprotein using a yeast peptide library.";
 RL Virology 200:246-255(1994).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA (BY SIMILARITY).
 CC EMBL: D10749; BAA01582.1; -.
 DR HSP: P27958; IHEI.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRp.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRp; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR PROSITE: PS50507; RDRP_POSITIVE; 1.
 DR PROSITE: PS50521; RDRP_VIRAL; 1.
 KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
 KW Hydrolase; Nonstructural protein; Polyprotein;
 KW RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3011 AA; 327112 MW; 97E9052C0250463B CRC64;

Query Match 87.3%; Score 900.5; DB 12; Length 3011;
 Best Local Similarity 85.8%; Pred. No. 3.2e-82;
 Matches 175; Conservative 8; Mismatches 12; Indels 9; Gaps 1;

QY 3 KGSVVIVGRIN-----LSGDTAYAAQOTRGECCOETSOTGRDKNQVEGEVQIVST 53
 DB 1005 RRGREILLGPADGMVSKGWRLLAPITAYAAQOTRGLGCIITSLTGRDKNQVEGEVQIVST 1064

QY 54 AAQTFLATCINGVCWTVYHGAGTRTIAISPKGPVIQMTYNDKDLVGPAPQGSRSILTPCT 113
 DB 1065 AAQTFLATCINGVCWTVYHGAGTRTIAISPKGPVIQMTYNDKDLVGPAPQGSRSILTPCT 1124

QY 114 CGSSDLYLVTRHADVIPVRRGDSRGLSPRISYLYKSSGGPLCPAGHAVGIFRAAV 173
 DB 1125 CGSSDLYLVTRHADVIPVRRGDSRGLSPRISYLYKSSGGPLCPAGHAVGIFRAAV 1184

```
QY 174 CTRGVAKAVDFIPVESLETTMRSP 197
DB 1185 CTRGVAKAVDFIPVESLETTMRSP 1208

RESULT 5
O36579 PRELIMINARY: PRT; 3011 AA.
AC O36579;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H77;
RX MEDLINE=97373636; PubMed=9228008;
RA Kolykhalov A.A., Agapov E.V., Blight K.J., Mihalik K., Feinstone S.M.,
RA Rice C.M.;
RT "Transmission of hepatitis C by intrahepatic inoculation with
RT transcribed RNA.";
RL Science 277:570-574(1997).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AF009606; AAB66324.1; -.
DR HSP; P27958; 1HE1.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_Core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NSI.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRP.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NSI; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR ProDom; PD186062; HCV_NSI; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS05057; RDRP_POSITIVE; 1.
DR PROSITE; PS05021; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA; 327182 MW; E2E0EE809C63C1B9 CRC64;

Query Match 86.98; Score 896.5; DB 12; Length 3011;
Best Local Similarity 84.88; Pred. No. 8.2e-82;
Matches 173; Conservative 10; Mismatches 12; Indels 9; Gaps 1;

QY 3 KKGVSIVGVGRIN-----LSGDTAYAAQTGEGCGQETSTGDRKNOVEGEVOIVST 53
DB 1005 RRGCEILLGADGVMSKGRLLAPITAYAAQTGGLGCIITSLTGRDKNQVEGEVOIVST 1064

RESULT 6
Q91RR8 PRELIMINARY: PRT; 181 AA.
AC Q91RR8;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PL.1Y;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Meyers D.L.;
RT "Genetic diversity and response to IFN of the NS3 Protease Gene from
RT Clinical Strains of the Hepatitis C Virus.";
RL EMBL; AF369235; AAK54560.1; -.
DR InterPro: IPR004109; HCV_NS3.
DR Pfam; PF02907; HCV_NS3; 1.
DR NON_TER 1 181
FT NON_TER 181
SQ SEQUENCE 181 AA; 19130 MW; 85D9186299B7C35 CRC64;

Query Match 86.88; Score 896; DB 12; Length 181;
Best Local Similarity 96.64; Pred. No. 2.5e-83;
Matches 172; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 19 TAYAAQTGEGCGQETSTGDRKNOVEGEVOIVSTAAQTFLATCINGVCWTVYHGAGT 78
DB 4 TAYAAQTGGLGCIITSLTGRDKNQVEGEVOIVSTAAQTFLATCINGVCWTVYHGAGT 63

QY 79 TASPQGVQIYNTNVKDLVGVPAQPGSRSLTPTCGSSDLYLVTRHADVIPVRRGDSR 138
DB 64 TASPQGVQIYNTNVKDLVGVPAQPGSRSLTPTCGSSDLYLVTRHADVIPVRRGDSR 123

QY 139 GSLLSPRISYLYKSGSGGLLCPCAGHAGVCIIPRAAVCTRGVAKAVDFIPVESLETTMRS 196
DB 124 GSLLSPRISYLYKSGSGGLLCPCAGHAGVCIIPRAAVCTRGVAKAVDFIPVESLETTMRS 181

RESULT 7
Q91RT5 PRELIMINARY: PRT; 181 AA.
AC Q91RT5;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PL.4;
```



```
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RL Clinical Strains of the Hepatitis C Virus.";
DR Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
EMBL: AF369218; AAK54543.1; -.
DR InterPro: IPR004109; HCV_NS3.
DR Pfam: PF02907; HCV_NS3.
KW Protease.
FT NON_TER 1
FT NON_TER 181
SQ SEQUENCE 181 AA; 19130 MW; 85D91869299B7C35 CRC64;

Query Match 86.8%; Score 896; DB 12; Length 181;
Best Local Similarity 96.1%; Pred. No. 2.5e-83;
Matches 172; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 19 TAYAAQTRGEGCGEETSQTRGKNOQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRT 78
DB 4 TAYAAQTRGLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRT 63
QY 79 IASPKGPVIQMTYNDKDLVGNPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRGDSR 138
DB 64 IASPKGPVIQMTYNDKDLVGNPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRGDSR 123
QY 139 GSLLSPRPISYLGSGGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMRS 196
DB 124 GSLLSPRPISYLGSGGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMRS 181

RESULT 8
QY1RR5
ID QY1RR5 PRELIMINARY; PRT; 181 AA.
AC QY1RR5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pt.30;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RL Clinical Strains of the Hepatitis C Virus.";
DR Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
EMBL: AF369238; AAK54563.1; -.
DR InterPro: IPR004109; HCV_NS3.
DR Pfam: PF02907; HCV_NS3.
KW Protease.
FT NON_TER 1
FT NON_TER 181
SQ SEQUENCE 181 AA; 19084 MW; 3B5E8161F2100A72 CRC64;

Query Match 86.6%; Score 894; DB 12; Length 181;
Best Local Similarity 96.1%; Pred. No. 4e-83;
Matches 171; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 19 TAYAAQTRGEGCGEETSQTRGKNOQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRT 78
DB 4 TAYAAQTRGLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRT 63
QY 79 IASPKGPVIQMTYNDKDLVGNPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRGDSR 138
DB 64 IASPKGPVIQMTYNDKDLVGNPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRGDSR 123
QY 139 GSLLSPRPISYLGSGGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMRS 196
DB 124 GSLLSPRPISYLGSGGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMRS 181
```

```
RESULT 9
QY1RR3
ID QY1RR3 PRELIMINARY; PRT; 181 AA.
AC QY1RR3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pt.4B;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RL Clinical Strains of the Hepatitis C Virus.";
DR Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
EMBL: AF369240; AAK54565.1; -.
DR InterPro: IPR004109; HCV_NS3.
DR Pfam: PF02907; HCV_NS3.
KW Protease.
FT NON_TER 1
FT NON_TER 181
SQ SEQUENCE 181 AA; 19115 MW; 5D85F88AD7AC1A11 CRC64;

Query Match 86.5%; Score 893; DB 12; Length 181;
Best Local Similarity 96.1%; Pred. No. 5.1e-83;
Matches 171; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 19 TAYAAQTRGEGCGEETSQTRGKNOQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRT 78
DB 4 TAYAAQTRGLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRT 63
QY 79 IASPKGPVIQMTYNDKDLVGNPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRGDSR 138
DB 64 IASPKGPVIQMTYNDKDLVGNPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRGDSR 123
QY 139 GSLLSPRPISYLGSGGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMRS 196
DB 124 GSLLSPRPISYLGSGGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMRS 181

RESULT 10
QY1RS1
ID QY1RS1 PRELIMINARY; PRT; 181 AA.
AC QY1RS1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pt.K;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RL Clinical Strains of the Hepatitis C Virus.";
DR Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
EMBL: AF369232; AAK54557.1; -.
DR InterPro: IPR004109; HCV_NS3.
DR Pfam: PF02907; HCV_NS3.
KW Protease.
FT NON_TER 1
FT NON_TER 181
SQ SEQUENCE 181 AA; 19114 MW; ABB9085B3ABA4E26 CRC64;

Query Match 86.5%; Score 893; DB 12; Length 181;
Best Local Similarity 96.1%; Pred. No. 5.1e-83;
```

```
Matches 171; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 19 TAYAAQTRGEGCQETSOTGRDKNQVEGEVOIVSTAAQTFLATCINGVCWTVYHGAGT 78
DB 4 TAYAAQTRGELGCIITSUTGRDKNQVEGEVOIVSTAAQTFLATCINGVCWTVYHGAGT 63
QY 79 IASPKGPVIQMYTNVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRGDSR 138
DB 64 IASPKGPVIQMYTNVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRGDSR 123
QY 139 GSLLSPRISYLKSSGGPPLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMRS 196
DB 124 GSLLSPRISYLKSSGGPPLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTMRS 181

RESULT 11
Q91R08
ID Q91R08 PRELIMINARY; PRT: 181 AA.
AC Q91R08;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Pt.5Z;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF369245; AAK54570.1; -
DR InterPro: IPR004109; HCV_NS3.
DR Pfam: PF02907; HCV_NS3; 1.
KW Protease.
FT NON_TER 1 1
FT NON_TER 181 181
FT SEQUENCE 181 AA; 19144 MW; C0C91F1E2EEB0B32 CRC64;

Query Match 86.5%; Score 893; DB 12; Length 181;
Best Local Similarity 96.1%; Pred. No. 5.le-83;
Matches 171; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 19 TAYAAQTRGEGCQETSOTGRDKNQVEGEVOIVSTAAQTFLATCINGVCWTVYHGAGT 78
DB 4 TAYAAQTRGELGCIITSUTGRDKNQVEGEVOIVSTAAQTFLATCINGVCWTVYHGAGT 63
QY 79 IASPKGPVIQMYTNVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRGDSR 138
DB 64 IASPKGPVIQMYTNVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRGDSR 123
QY 139 GSLLSPRISYLKSSGGPPLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMRS 196
DB 124 GSLLSPRISYLKSSGGPPLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTMRS 181

RESULT 12
Q91RT1
ID Q91RT1 PRELIMINARY; PRT: 181 AA.
AC Q91RT1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN-Pt.161;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF369222; AAK54547.1; -
DR InterPro: IPR004109; HCV_NS3.
DR Pfam: PF02907; HCV_NS3; 1.
KW Protease.
FT NON_TER 1 1
FT NON_TER 181 181
FT SEQUENCE 181 AA; 19114 MW; ABB90B5B3ABA4E26 CRC64;

Query Match 86.5%; Score 893; DB 12; Length 181;
Best Local Similarity 96.1%; Pred. No. 5.le-83;
Matches 171; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 19 TAYAAQTRGEGCQETSOTGRDKNQVEGEVOIVSTAAQTFLATCINGVCWTVYHGAGT 78
DB 4 TAYAAQTRGELGCIITSUTGRDKNQVEGEVOIVSTAAQTFLATCINGVCWTVYHGAGT 63
QY 79 IASPKGPVIQMYTNVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRGDSR 138
DB 64 IASPKGPVIQMYTNVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRGDSR 123
QY 139 GSLLSPRISYLKSSGGPPLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMRS 196
DB 124 GSLLSPRISYLKSSGGPPLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTMRS 181

RESULT 13
Q9ELS8
ID Q9ELS8 PRELIMINARY; PRT: 3011 AA.
AC Q9ELS8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-colonel;
RA Desai S.M., Devare S., Yamaguchi J.;
RT "Hepatitis C Virus.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: AF290978; AAG02099.1; -
DR HSP: P27958; IHEI
DR InterPro: IPR000345; CytC_heme_bind.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRp.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
```

DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; Helicase_C; 1.
 DR Pfam: PF00998; Viral_RDRP; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR PROSITE: PS00190; CYTOCHROME_C; 1.
 DR PROSITE: PS05057; RDRP_POSITIVE; 1.
 DR PROSITE: PS0521; RDRP_VIRAL; 1.
 KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
 KW Hydrolase; Nonstructural protein; Polyprotein;
 KW RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3011 AA; 327107 MW; A6BECF5A3B3EE13F CRC64;

Query Match 86.5%; Score 892.5; DB 12; Length 3011;
 Best Local Similarity 84.3%; Pred. No. 2.1e-81;
 Matches 172; Conservative 11; Mismatches 12; Indels 9; Gaps 1;

QY 3 KKGSVVIVGRIN-----LSGDTAYAAQTRGEGCOETSGTGRDKNOVEGEVQIVST 53
 DB 1005 RRGQELLGPADGMVSKGWRLLAPITAYAAQTRGLLGCITSLTGRDKNOVEGEVQIVST 1064

QY 54 AAQTFLATCINGVCVTVYHGAGTRTIAAPKGPVIQMTNVKDLVGVWPAPOGSRSLTPTCT 113
 DB 1065 ATOTFLATCINGVCVTVYHGAGTRTIAAPKGPVIQMTNVKDLVGVWPAPOGSRSLTPTCT 1124

QY 114 CGSSDLVYTRHADVPIVRRGDSGSLSPRISYLGSSGGPPLCPAGHAGVIFRAAV 173
 DB 1125 CGSSDLVYTRHADVPIVRRGDSGSLSPRISYLGSSGGPPLCPAGHAGVIFRAAV 1184

QY 174 CTRGVAKAVDFIPVESLETMRSP 197
 DB 1185 CTRGVAKAVDFIPVENLETMRSP 1208

RESULT 14

Q91RR6 PRELIMINARY; PRT; 181 AA.
 AC Q91RR6;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE NS3 protease (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP STRAIN=Pt.3T;
 RC Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
 RA "Genetic Diversity and Response to IFN of the NS3 Protease Gene from
 RT Clinical Strains of the Hepatitis C Virus."
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF369237; AAK54562.1;
 DR InterPro: IP0004109; HCV_NS3.
 DR Pfam: PF02907; HCV_NS3; 1.
 KW Protease.
 FT NON_TER 1
 FT NON_TER 181
 SQ SEQUENCE 181 AA; 19101 MW; 614ADA8B0F33CCAF CRC64;

Query Match 86.3%; Score 891; DB 12; Length 181;
 Best Local Similarity 95.5%; Pred. No. 8.2e-83;
 Matches 170; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 19 TAYAAQTRGEGCOETSGTGRDKNOVEGEVQIVSTAAQTFLATCINGVCVTVYHGAGTRT 78
 DB 4 TAYAAQTRGLLGCITSLTGRDKNOVEGEVQIVSTAAQTFLATCINGVCVTVYHGAGTRT 63

QY 79 IASPKGPVIQMTNVKDLVGVWPAPOGSRSLTPTCTCGSSDLVYTRHADVPIVRRGDSR 138
 DB 64 IASPKGPVIQMTNVKDLVGVWPAPOGSRSLTPTCTCGSSDLVYTRHADVPIVRRGDSR 123

QY 139 GSLLSPRPISYLGSSGGPPLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETMR 196
 DB 124 GSLLSPRPISYLGSSGGPPLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETMR 181

RESULT 15
 Q91RS9 PRELIMINARY; PRT; 181 AA.
 ID Q91RS9;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE NS3 protease (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Pt.174;
 RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
 RT "Genetic Diversity and Response to IFN of the NS3 Protease Gene from
 RL Clinical Strains of the Hepatitis C Virus."
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF369224; AAK54549.1;
 DR InterPro: IP0004109; HCV_NS3.
 DR Pfam: PF02907; HCV_NS3; 1.
 KW Protease.
 FT NON_TER 1
 FT NON_TER 181
 SQ SEQUENCE 181 AA; 19131 MW; 8BD7FC2769DBD635 CRC64;

Query Match 86.3%; Score 891; DB 12; Length 181;
 Best Local Similarity 96.1%; Pred. No. 8.2e-83;
 Matches 171; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 19 TAYAAQTRGEGCOETSGTGRDKNOVEGEVQIVSTAAQTFLATCINGVCVTVYHGAGTRT 78
 DB 4 TAYAAQTRGLLGCITSLTGRDKNOVEGEVQIVSTAAQTFLATCINGVCVTVYHGAGTRT 63

QY 79 IASPKGPVIQMTNVKDLVGVWPAPOGSRSLTPTCTCGSSDLVYTRHADVPIVRRGDSR 138
 DB 64 IASPKGPVIQMTNVKDLVGVWPAPOGSRSLTPTCTCGSSDLVYTRHADVPIVRRGDSR 123

QY 139 GSLLSPRPISYLGSSGGPPLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETMR 196
 DB 124 GSLLSPRPISYLGSSGGPPLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETMR 181

Search completed: August 30, 2003, 19:18:19
 Job time : 38.5921 secs

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	884.5	85.7	3011	1	GNMVC3	genome polyprotein
2	883.5	85.6	3011	1	S40770	genome polyprotein
3	878.5	85.1	3011	1	GNMVCB	genome polyprotein
4	867.5	84.1	3010	1	GNMVTW	genome polyprotein
5	857.5	83.1	3010	1	A45573	genome polyprotein
6	853.5	82.7	3010	1	GNMVC7	genome polyprotein
7	853.5	82.7	3010	1	GNMVCJ	genome polyprotein
8	842.5	81.6	3010	1	S18030	genome polyprotein
9	760.5	73.7	3014	1	JC5620	genome polyprotein
10	677	65.6	3033	1	QJ1303	genome polyprotein
11	675	65.4	3033	1	GNMWJ8	genome polyprotein
12	257	24.9	3005	2	T08841	polyprotein - dour
13	251	24.3	2970	2	T08839	polyprotein - marm
14	88.5	8.6	590	2	B81104	nitrate/nitrite se
15	88.5	8.6	590	2	C81911	nitrate/nitrite se
16	84.5	8.2	492	2	AH1030	probable exported
17	84.5	8.2	495	2	B71360	hypothetical prote
18	81	7.8	452	2	I39383	angio-associated m
19	79	7.7	477	2	E75392	hypothetical prote
20	78	7.6	915	2	F81196	transferrin-bindin
21	78	7.6	1615	2	JE0372	low density lipopr
22	77	7.5	322	2	D87603	glycosyl transfera
23	77	7.5	2508	2	S61441	surface-associated
24	76.5	7.4	264	2	I38136	chymotrypsin-like
25	76.5	7.4	323	3	PR1JHD	proteinase (EC 3.4
26	76.5	7.4	333	1	B53308	mosa protein - Rhi
27	76.5	7.4	398	2	B71284	probable periplasm
28	76.5	7.4	716	2	G83612	hypothetical prote
29	76	7.4	911	2	JN0821	transferrin-bindin

```

Db      1005  RRGREILLGPADGMVSKGWRLLAPITAYAAQOTRGLLGCIIITSLTGRDKNQVEGEVQIVST 1064
QY      54  ATOTFLATCINGVCWTVYHGAGTRTIASPKGPVTOMYTNVDKDLVGMQAPQGSRLTPTCT 113
      1  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1065  AAQTFLATCINGVCWTVYHGAGTRTIASPKGPVTOMYTNVDQDLVGMQAPQGSRLTPTCT 1124
QY      114  CGSSDLVLTTRHADVIPVRRGDSRGLSPRPISYLGSSGGPGLLCPAGHAGVIFRAAV 173
      1  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1125  CGSSDLVLTTRHADVIPVRRGDSRGLSPRPISYLGSSGGPGLLCPAGHAGVIFRAAV 1184
QY      174  CTRGVAKAVDFIPVESLETTMRSP 197
      1  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1185  CTRGVAKAVDFIPVENLETTMRSP 1208
      1  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 2
S40770
genome polyprotein - hepatitis C virus
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain H)
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C:Accession: S40770; PC1285
R:Okamoto, H.
submitted to the EMBL Data Library, March 1992
A:Reference number: S40770
A:Accession: S40770
A:Molecule type: genomic RNA
A:Residues: 1-3011 <OK>
A:Cross-references: EMBL:D10749; NID:g221586; PIDN:BAA01582.1; PID:g221587
R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda,
Jpn. J. Exp. Med. 60, 167-177, 1990
A:Title: The 5'-terminal sequence of the hepatitis C virus genome.
A:Reference number: PC1284; MUID:91013116; PMID:2170712
A:Accession: PC1285
A:Molecule type: genomic RNA
A:Residues: 1-513 <OK>
A:Cross-references: GB:D00831; NID:g221511; PIDN:BAA00705.1; PID:g221512
A:Experimental source: isolate HC-J1
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polyprotein
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 85.6%; Score 883.5; DB 1; Length 3011;
Best Local Similarity 84.3%; Pred. No. 8.8e-73;
Matches 172; Conservative 8; Mismatches 15; Indels 9; Gaps 1;

QY      3  KKGSVVIVGRIN-----LSGDTAYAAQOTRGECCOETSGTGRDKNQVEGEVQIVST 53
      1  || :|| : ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1005  RKGREILLGPADGMVSKGWRLLAPITAYAAQOTRGLLGCIIITSLTGRDKNQVEGEVQIVST 1064
QY      54  ATOTFLATCINGVCWTVYHGAGTRTIASPKGPVTOMYTNVDKDLVGMQAPQGSRLTPTCT 113
      1  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1065  AAQTFLATCINGVCWTVYHGAGTRTIASPKGPVTOMYTNVDQDLVGMQAPQGSRLTPTCT 1124
QY      114  CGSSDLVLTTRHADVIPVRRGDSRGLSPRPISYLGSSGGPGLLCPAGHAGVIFRAAV 173
      1  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1125  CGSSDLVLTTRHADVIPVRRGDSRGLSPRPISYLGSSGGPGLLCPAGHAGVIFRAAV 1184
QY      174  CTRGVAKAVDFIPVESLETTMRSP 197
      1  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1185  CTRGVAKAVDFIPVENLETTMRSP 1208
      1  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 4
S40770
genome polyprotein - hepatitis C virus (strain Taiwan)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain Taiwan)
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C:Accession: A40244
R:Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.
Virology 188, 102-113, 1992

```

RESULT 3

GNVYCH

```

genome polyprotein - hepatitis C virus (strain H)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain H)
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C:Accession: A36814; A41546
R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
submitted to GenBank, July 1992
A:Description: Genomic structure of the human prototype strain H of hepatitis C virus
A:Reference number: A41546; MUID:92052256; PMID:1658800
A:Contents: annotation
A:Note: neither amino acid nor nucleotide sequence is given
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruct
F:1-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polyprotein
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F:196, 209, 234, 305, 325, 417, 423, 430, 448, 476, 532, 540, 556, 576, 623, 645, 1213, 1255, 2041, 2240

```

Query Match

85.1%; Score 878.5; DB 1; Length 3011;

Best Local Similarity 83.3%; Pred. No. 2.5e-72;

Matches 170; Conservative 10; Mismatches 15; Indels 9; Gaps 1;

QY 3 KKGSVVIVGRIN-----LSGDTAYAAQOTRGECCOETSGTGRDKNQVEGEVQIVST 53

:|| :|| : ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1005 RKGREILLGPADGMVSKGWRLLAPITAYAAQOTRGLLGCIIITSLTGRDKNQVEGEVQIVST 1064

QY 54 ATOTFLATCINGVCWTVYHGAGTRTIASPKGPVTOMYTNVDKDLVGMQAPQGSRLTPTCT 113

:|| :|| : ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1065 ATOTFLATCINGVCWTVYHGAGTRTIASPKGPVTOMYTNVDQDLVGMQAPQGSRLTPTCT 1124

QY 114 CGSSDLVLTTRHADVIPVRRGDSRGLSPRPISYLGSSGGPGLLCPAGHAGVIFRAAV 173

:|| :|| : ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1125 CGSSDLVLTTRHADVIPVRRGDSRGLSPRPISYLGSSGGPGLLCPAGHAGVIFRAAV 1184

QY 174 CTRGVAKAVDFIPVESLETTMRSP 197

:|| :|| : ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1185 CTRGVAKAVDFIPVENLETTMRSP 1208

:|| :|| : ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 4

GNVYCH

```

genome polyprotein - hepatitis C virus (strain Taiwan)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain Taiwan)
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C:Accession: A40244
R:Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.
Virology 188, 102-113, 1992

```

A:Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the
 A:Reference number: A40244; MUID:92230206; PMID:1314449
 A:Accession: A40244
 A:Molecule type: genomic RNA
 A:Residues: 1-3010 <CHE>
 A:Cross-references: GB:M84754
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
 F:1-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <MEE>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: hepatitis C virus genome polyprotein
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B
 F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,207

Query Match 84.1%; Score 867.5; DB 1; Length 3010;
 Best Local Similarity 79.9%; Pred. No. 2.6e-71;
 Matches 163; Conservative 18; Mismatches 14; Indels 9; Gaps 1;

QY 3 KGSVVIVGRIN-----LSGDTAYAAQOTRGECCOETSOTGRDNQVEGEVQIVST 53
 DB 1005 RRGEILLGPADSLRGWRLAPITAYAAQOTRGLFGCIITSLTRDNQVEGEVQIVST 1064

QY 54 ATQTFLATCINGVCWTVYHGAGTRTASPKGPVTOMYTNVDKDLVQWAPQGSRSITPCT 113
 DB 1065 ATQSFLATCINGVCWTVYHGAGSKTLAGPKGPIOMYTNVDQDLVGHAPQGSRSITPCT 1124

QY 114 CGSSDLVLTTRHADVIPVRRGDSRGLSPRISYLYKSGSGGLPCLCPAGHAGVIFRAAV 173
 DB 1125 CGSSDLVLTTRHADVIPVRRGDSRGLSPRISYLYKSGSGGLPCLCPGSHAGVIFRAAV 1184

QY 174 CTRGAKAVDFIPVESLETTMRSP 197
 DB 1185 CTRGAKAVDFIPVESMETTMRSP 1208

RESULT 5
 A45573
 genome polyprotein - hepatitis C virus (strain JT)
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein
 Protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
 C:Accession: A45573
 R:Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata,
 Virus Res. 23, 39-53, 1992
 A:Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: s
 A:Reference number: A45573; MUID:92295714; PMID:1318627
 A:Accession: A45573
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-3010 <TAN>
 A:Cross-references: GB:D11168; GB:D01171; MID:g221612; PIDN:BAA01943.1; PID:g221613
 A:Experimental source: HCV-JT
 A:Note: sequence extracted from NCBI backbone (NCBI:106206, NCBI:106207)
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
 F:2-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <MEE>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: hepatitis C virus genome polyprotein
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B
 F:1316-1319/Region: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 83.1%; Score 857.5; DB 1; Length 3010;
 Best Local Similarity 78.4%; Pred. No. 2.2e-70;
 Matches 160; Conservative 20; Mismatches 15; Indels 9; Gaps 1;

QY 3 KGSVVIVGRIN-----LSGDTAYAAQOTRGECCOETSOTGRDNQVEGEVQIVST 53
 DB 1005 RRGEILLGPADSLRGWRLAPITAYAAQOTRGLGCIITSLTRDNQVEGEVQIVST 1064

QY 54 ATQTFLATCINGVCWTVYHGAGTRTASPKGPVTOMYTNVDKDLVQWAPQGSRSITPCT 113
 DB 1065 ATQSFLATCINGVCWTVYHGAGSKTLAGPKGPIOMYTNVDQDLVGHAPQGSRSITPCT 1124

QY 114 CGSSDLVLTTRHADVIPVRRGDSRGLSPRISYLYKSGSGGLPCLCPAGHAGVIFRAAV 173
 DB 1125 CGSSDLVLTTRHADVIPVRRGDSRGLSPRISYLYKSGSGGLPCLCPGSHAGVIFRAAV 1184

QY 174 CTRGAKAVDFIPVESLETTMRSP 197
 DB 1185 CTRGAKAVDFIPVESMETTMRSP 1208

RESULT 6
 GNVVTC
 genome polyprotein - hepatitis C virus
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein
 Protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 19-Jan-2001
 C:Accession: A38465
 R:Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, I
 J. Virol. 65, 1105-1113, 1991
 A:Title: Structure and organization of the hepatitis C virus genome isolated from hur
 A:Reference number: A38465; MUID:91140698; PMID:1847440
 A:Accession: A38465
 A:Molecule type: genomic RNA
 A:Residues: 1-3010 <TAK>
 A:Cross-references: EMBL:M58335; MID:g329770; PIDN:AAA72945.1; PID:g329771
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruct
 F:2-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <MEE>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: hepatitis C virus genome polyprotein
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B
 F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,207

Query Match 82.7%; Score 853.5; DB 1; Length 3010;
 Best Local Similarity 77.9%; Pred. No. 5.1e-70;
 Matches 159; Conservative 21; Mismatches 15; Indels 9; Gaps 1;

QY 3 KGSVVIVGRIN-----LSGDTAYAAQOTRGECCOETSOTGRDNQVEGEVQIVST 53
 DB 1005 RRGEILLGPADSLRGWRLAPITAYAAQOTRGLGCIITSLTRDNQVEGEVQIVST 1064

QY 54 ATQTFLATCINGVCWTVYHGAGTRTASPKGPVTOMYTNVDKDLVQWAPQGSRSITPCT 113
 DB 1065 ATQSFLATCINGVCWTVYHGAGSKTLAGPKGPIOMYTNVDQDLVGHAPQGSRSITPCT 1124

QY 114 CGSSDLVLTTRHADVIPVRRGDSRGLSPRISYLYKSGSGGLPCLCPAGHAGVIFRAAV 173
 DB 1125 CGSSDLVLTTRHADVIPVRRGDSRGLSPRISYLYKSGSGGLPCLCPGSHAGVIFRAAV 1184

A:Accession: J05620
A:Molecule type: mRNA
A:Residues: 1-3014 <CHA>
A:Cross-references: GB:Y13184
A:Experimental source: isolate 5a, which predominates in South Africa
A:Note: the translation of the nucleotide sequence is not complete in this paper
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serine
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:384-408/Region: hypervariable #status predicted
F:390-733/Product: nonstructural protein NS1 #status predicted <NS1>
F:731-1007/Product: nonstructural protein NS2 #status predicted <NS2>
F:1008-1616/Product: hepatitis C virus genome polyprotein NS4 #status predicted <NS4>
F:1231-1238/Region: nucleotide-binding motif A (P-loop)
F:1313-1318/Region: nucleotide-binding motif B
F:1317-1320/Region: DEXH motif
F:1617-1863/Product: nonstructural protein NS4a #status predicted <NS4a>
F:1864-2014/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>
F:2210-2249/Region: interferon sensitivity determining #status predicted

Query Match 73.7%; Score 760.5; DB 1; Length 3014;
Best Local Similarity 69.8%; Pred. No. 1.9e-61;
Matches 142; Conservative 24; Mismatches 29; Indels 9; Gaps 1;

QY 3 KKGSVVIVGRIN-----LSGDTAYAQTRGEGCQETSGTGRKKNQVEGQIVST 53
DB 1006 RRGREIFLGPADDIKTSGHRLAPITAYAQTRGVGLAIVLSLTGRKNEAGEVQFLST 1065

QY 54 ATGTFLATCINGVCVTVYHGAGTRTASPKGPVTQMYTNVDKDLVGWQAPQSGRSRSLTPT 113
DB 1066 ATGTFLGICINGVMTLFGAGSKTLAGPKGPVQVQMYTNVDKDLVGWQAPQSGRSRSLTPT 1125

QY 114 CGSSDLYLTVTRADYIPVRRGRSGSLSPRISVLKSSGGPILCPAGHAGVIFRAAV 173
DB 1126 CGSADLYLTVTRADYIPARRRGDTRASLSRPRISVLKSSGGPILCPAGHAGVIFRAAV 1185

QY 174 CTGKGVAKAVDFIPVESLETTMRSP 197
DB 1186 CTGKGVAKALEFVVENLETTMRSP 1209

RESULT 10
JQ1303
genome polyprotein - hepatitis C virus (isolate HC-J6)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome having poor homology to
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 17-Nov-2000
C:Accession: JQ1303
R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Kurai, K.; Lizuka, H.; Machida, A.; Miyakawa, Y.
J. Gen. Virol. 72, 2697-2704, 1991
A:Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human
A:Reference number: JQ1303; MUID:9204440; PMID:1658196
A:Accession: JQ1303
A:Molecule type: genomic RNA
A:Residues: 1-3033 <OKA>
A:Cross-references: GB:D00944; NID:g221650; PIDN:BAA00792.1; PID:g221651
A:Experimental source: isolate HC-J6 from a Japanese individual
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; transmembrane
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-733/Product: nonstructural protein NS1 #status predicted <NS1>
F:731-1007/Product: nonstructural protein NS2 #status predicted <NS2>
F:1011-1619/Product: nonstructural protein NS4a #status predicted <NS4a>
F:1316-1321/Region: nucleotide-binding motif A
F:1317-1320/Region: DEXH motif
F:1617-1863/Product: nonstructural protein NS4b #status predicted <NS4b>
F:1864-2014/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 65.4%; Score 675; DB 1; Length 3033;
Best Local Similarity 69.3%; Pred. No. 1.5e-53;
Matches 124; Conservative 24; Mismatches 31; Indels 0; Gaps 0;

F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,305,325,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,203

Query Match 65.6%; Score 677; DB 1; Length 3033;
Best Local Similarity 68.7%; Pred. No. 9.8e-54;
Matches 123; Conservative 26; Mismatches 30; Indels 0; Gaps 0;

QY 19 TAYAQTRGEGCQETSGTGRKKNQVEGQIVSTATOTFLATCINGVCVTVYHGAGTRT 78
DB 1034 TAYAQTRGLLGTIVSMTGRDTEQAGEIQVLSVTQSFGLTTSIGVLTVYHGAGNKT 1093

QY 79 IASPKGPVTQMYTNVDKDLVGWQAPQSGRSRSLTPTCGSSDLYLTVTRADYIPVRRGRSGR 138
DB 1094 LAGSRGPVTQMYSSAEGDLVGPSPPTGKSLPECTCGAVDLYLTVTRADYIPARRGDKR 1153

QY 139 GSSLSPRISVLKSSGGPILCPAGHAGVIFRAAVCTGKGVAKAVDFIPVESLETTMRSP 197
DB 1154 GALLSPRLSTLKGSSGGPVLCPRGHAGVIFRAAVCTGKGVAKSDFIPVETLDTVTRSP 1212

RESULT 11
GNWJ38
genome polyprotein - hepatitis C virus (strain HC-38)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome having poor homology to
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C:Accession: A40250; P00397; P00559
R:Okamoto, H.; Kurai, K.; Okada, S.I.; Yamamoto, K.; Lizuka, H.; Tanaka, T.; Fukuda, Y.
Virology 186, 331-341, 1992
A:Title: Full-length sequence of a hepatitis C virus genome having poor homology to
A:Reference number: A40250; MUID:92230232; PMID:1314459
A:Accession: A40250
A:Molecule type: genomic RNA
A:Residues: 1-3033 <OKA>
A:Cross-references: GB:D10988; GB:D01221; NID:g221608; PIDN:BAA01761.1; PID:g221609
R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, J. Gen. Virol. 73, 1131-1141, 1992
A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship
A:Reference number: P00393; MUID:92268871; PMID:1316939
A:Accession: P00397
A:Molecule type: genomic RNA
A:Residues: 2678-2754 <CHA>
A:Cross-references: DDBJ:D10134
A:Experimental source: isolate E-b12
R:Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Mori, S.; Hijikata, M.; Shimotoh
Biochem. Biophys. Res. Commun. 181, 279-285, 1991
A:Title: Distribution of plural HCV types in Japan.
A:Reference number: P00554; MUID:92068204; PMID:1720309
A:Accession: P00559
A:Molecule type: mRNA
A:Residues: 2678-2729 <KAT>
A:Cross-references: GB:D00518; NID:g221523; PIDN:BAA01418.1; PID:g221524
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruc
F:116-191/Product: capsid protein C #status predicted <CPC>
F:192-389/Product: major envelope protein E #status predicted <EPM>
F:390-733/Product: nonstructural protein NS1 #status predicted <NS1>
F:731-1010/Product: nonstructural protein NS2 #status predicted <NS2>
F:1011-1619/Product: hepatitis C virus genome polyprotein NS4a #status predicted <NS4a>
F:1316-1321/Region: nucleotide-binding motif A (P-loop)
F:1320-1323/Region: DEXH motif
F:1620-1866/Product: nonstructural protein NS4b #status predicted <NS4b>
F:1867-2017/Product: nonstructural protein NS5 #status predicted <NS5>
F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,233,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,203

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 30, 2003, 19:18:33 ; Search time 2560.57 Seconds
(without alignments)
3147.423 Million cell updates/sec

Title: US-09-965-594-14

Perfect score: 1032

Sequence: 1 MKKGSVVIVGRINLSGDTA.....VAKAVDFIPVESLETTMRSP 197

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-O=/cn2.1/USPTO_spool/US09965594/runat_29082003_151919_28310/app_query.fasta_1.2872
-DB=GenEmbl -ORFPT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09965594.ecgn_1.1.14686.e_runat_29082003_151919_28310 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORE=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6
-FCGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.scs:*
12: gb.sy:*
13: gb.un:*
14: gb.vi:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.mu:*
20: em.om:*
21: em.or:*
22: em.ov:*
23: em.pat:*
24: em.ph:*
25: em.pl:*
26: em.ro:*
27: em.sts:*
28: em.un:*

29: em.vi:*
30: em.htg_hum:*
31: em.htg_inv:*
32: em.htg_other:*
33: em.htg_mus:*
34: em.htg_pln:*
35: em.htg_rod:*
36: em.htg_mam:*
37: em.htg_vrt:*
38: em.sy:*
39: em.htgo_hum:*
40: em.htgo_mus:*
41: em.htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Query Length	DB ID	Description
1	930.5	90.2	12734	6	AR179057 Sequence
2	910.5	88.2	12734	14	AF268278 Pestiviru
3	901.5	87.4	5360	6	AR118686 Sequence
4	901.5	87.4	5360	6	I06434 Sequence 48
5	901.5	87.4	5360	6	I09328 Sequence 8
6	901.5	87.4	6785	6	AR118692 Sequence
7	901.5	87.4	6785	6	I06440 Sequence 54
8	901.5	87.4	6785	6	I09329 Sequence 10
9	901.5	87.4	7310	6	AR118696 Sequence
10	901.5	87.4	7310	6	I09331 Sequence 15
11	901.5	87.4	7310	14	HPCPOLYP M32084 Hepatitis C
12	901.5	87.4	8316	6	AR118703 Sequence
13	901.5	87.4	8987	6	AR118728 Sequence
14	901.5	87.4	9185	6	AR118722 Sequence
15	901.5	87.4	9185	6	AR118723 Sequence
16	901.5	87.4	9185	6	BD091382 Sequence
17	901.5	87.4	9185	6	I08294 Sequence 1
18	901.5	87.4	9379	6	AR166930 Sequence
19	901.5	87.4	9379	6	AR301300 Sequence
20	901.5	87.4	9401	6	AR176483 Sequence
21	901.5	87.4	9401	6	BD080334 Sequence
22	901.5	87.4	9401	6	E66593 Hepatitis C
23	901.5	87.4	9401	6	I71894 Sequence 9
24	901.5	87.4	9401	6	I81885 Sequence 9
25	901.5	87.4	9401	14	HPGPLYPRE M62321 Hepatitis C
26	901.5	87.4	9609	12	AF387805 Synthetic
27	901.5	87.4	9609	12	AF387808 Synthetic
28	901.5	87.4	9618	14	AF271632 Hepatitis
29	901.5	87.4	9646	12	AF387806 Synthetic
30	901.5	87.4	9693	12	AF387807 Synthetic
31	900.5	87.3	9502	6	E08263 gRNA of Hep
32	900.5	87.3	9502	6	E08264 cDNA of Hep
33	900.5	87.3	9502	14	HPCHCJ1 D10749 Hepatitis C
34	899	87.1	2058	6	AX395309 Sequence
35	899	87.1	2058	6	AX454818 Sequence
36	899	87.1	8157	6	AR127810 Sequence
37	899	87.1	8157	6	BD081911 Sequence
38	898.5	87.1	9424	14	AF511948 Hepatitis
39	897	86.9	1932	6	AR127809 Sequence
40	897	86.9	1932	6	BD081910 Hepatitis
41	896.5	86.9	9646	6	AR110828 Sequence
42	896.5	86.9	9646	6	BD069982 Functiona
43	896.5	86.9	9646	14	AF009606 Hepatitis
44	896.5	86.9	12980	6	AR110831 Sequence
45	896.5	86.9	12980	6	BD069985 Functiona

ALIGNMENTS

RESULT 1

AR179057 LOCUS AR179057 12734 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6326137.
ACCESSION AR179057
VERSION AR179057.1 GI:20220612
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 12734)
AUTHORS Hong, Z., Lai, V.C.H. and Lau, J.Y.N.
TITLE Hepatitis C virus protease-dependent chimeric pestivirus
JOURNAL Patent: US 6326137-A 1 04-DEC-2001;
FEATURES
Location/Qualifiers
source 1..12734
BASE COUNT 4032 a 2604 c 3295 g 2803 t
ORIGIN
Alignment Scores:
Pred. No.: 5, 27e-65 Length: 12734
Score: 930.50 Matches: 181
Percent Similarity: 94.87% Conservative: 4
Best Local Similarity: 92.82% Mismatches: 7
Query Match: 90.16% Indels: 3
DB: 6 Gaps: 1
US-09-965-594-14 (1-197) x AR179057 (1-12734)
QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
DB 413 GGTAGTGTGTATATGTTGGTAGAATGTTTATCTGCTAGTGTAGTACGACGGGTAC 472
QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGlnThrSerGlnThrGlyArgAspLys 41
DB 473 GCCACAGACGACGAGCCCTCCTAGGCTGTAGATCACCAGTCTGACTGCCGGGACAA 532
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAlaThr 61
DB 533 AACCAAGTGGAGGCTGAGTCCAGATCGTCTCACTGCTACCAAACTTCTCCGGCAACG 592
QY 62 CysIleAsnGlyValCysTyrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
DB 593 TGCAATCAATGGGTGTGCTGTGACTGTCTACACGGGCGGACGACGAGACCATCGCATCA 652
QY 82 ProLysGlyProValIleGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrPro 101
DB 653 CCCAAGGCTCCTGCATCCACAGATGATACCAATGTGCACCAAGACCTGTGGGCTGCC 712
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
DB 713 GCTCCTCAAGGTTCCCGCTCAITGACACCTGTCACCTGCGGCTCCTCGGACCTTTACCTG 772
QY 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141
DB 773 GTTAGGACGACGCGGAGCTCATTCCTCGCGCGGAGGTGATAGCAGGGGTAGCGCTG 832
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
DB 833 CTTTCGCCCGCGCCCATTTCTACTCTAAAGGCTCCTCGGGGGGTGCGCTGTGTGCC 892
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181
DB 893 GCGGACACGCGGTGGGCTATTTCAGGCGCGGCTGTGCACCGGTGGAGTGGCCAGGCG 952
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196
DB 953 GTGACTTTATCCTGTGGAGAACCTAGACAGAACCATGAGATCC 997
RESULT 2
LOCUS AF268278 12734 bp RNA linear VRL 12-JUL-2000
DEFINITION Pestivirus type 1, complete genome.

AF268278
AF268278.1 GI:9049956
Pestivirus type 1
Pestivirus type 1
Viruses: ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
pestivirus 1 (bases 1 to 12734)
Lai, V.C., Zhong, W., Skelton, A., Ingravallo, P., Vassilev, V.,
Donis, R.O., Hong, Z. and Lau, J.Y.
Generation and characterization of a hepatitis C virus NS3
protease-dependent bovine viral diarrhoea virus
J. Virol. 74 (14), 6339-6347 (2000)
20323484
10864644
2 (bases 1 to 12734)
Lai, V.C.H. and Hong, Z.
Direct Submission
Submitted (16-MAY-2000) Antiviral Therapy, Schering-Plough Research
Institute, 2015 Galloping Hill Road, Kenilworth, NJ 07033-0539, USA
Location/Qualifiers
1..12734
/organism="Pestivirus type 1"
/mol_type="genomic RNA"
/db_xref="taxon:11099"
1..385
386..12508
/codon_start=1
/product="polyprotein"
/protein_id="AAF82566.1"
/db_xref="GI:9049957"
/translation="MELTNEGSGSVVIVGRIVLSGSGSITACAQOTRGLLGCKITSL
TRGRNKGVEQIVSTATQTFLATCINGVCTVYHGAGTRTASGKPVQIYNTVD
QDLVWPAQGSRSRLTPTCGSDLYLVTRHANVIVPVRGRGDSRGLSPRLSYLKG
SSGGLPLCPAGHAGVLEFRAVCTRGKAVDFIPVENLETTTBSGSGATEDVVCSSM
SYSDTBEGATKKTKQPDRLERGMKIVPKSEKSKTTPDPATIVBEGVKVQVRKK
GKTSKNTDGLYHNKRPQESRKLKALLAIAIVLDFQVTMGNTITOMNLQNG
TEGORAMFORGVNRSUHGWEKICTGVSPSHLATDIELKTIHMGDASKTNTCCR
LORHEMKNHCWNIENPWLNVNRTQANLTGEPQPCAVTCRDDASDLNVVQA
RDSPTLTGCKKGNFAGILMRGPCNFIAASDLVLFKEHERISMFQDTTLVYDGL
TNSLGAORGTAKLTWLGKLGKLENKSKTNFAYAAAPYCDVSDRFGIYWT
KNCTPACPLKNTKIVGPKPDTNAEDCKILHEMGGHLSVLLLSLVLPDPAETASV
MYLILHSIPQSHVDVMDCKTQLNLVELTADVIFGVSVMNLGKWKVIRPNWMPYET
TVVLAPEVSQVVKLVLRALRDLTRINAAATTAFLCLVKIVRGQVQVQILMLLIT
VGQHLCKPEFYAIAKDERIGQGAELTTTWKESPGKLEDMVIAWEDGLM
YLOKCTRETRYLAITRALPTSVFKLFDGRQEDVDMNDFEGLPCDAKIV
RGKNTLLINGPQVPCIGTGTCTSCFNMDTLATTVTVTRRSKPPHPHQGIT
OKNLGDLHNCILGNNWTCVPGDOLLYKGSIESCKWGYOFKESGLPHYPIGCKIL
ENETGYRLVDSSTSCNRGVAIVPGTLLKCKTGKTKVQVIAMDTKLGMPRCRVEILSS
EGPKETATNTYTKLKYFEPDRDSYFQYMLKGEYQWEDLEVTDRHEDYFASI
LVVVALLGGRVILWLLVTHVLSQKALGTYGSGEVMAMGMLLTHNLEVTYFILL
LYLLRESVKKNVLLTHILLVHPKISVILLIMGVVKAQSGGQYELGRIDLQFT
TVVLIVIGLIARDDPTIPLVTIINALRVTELTTHQPGVDIAVAVMTITLLMYSYVTR
YFRYKMQCILSLVSGVFLRSILYLGRIEMPEVTIPNRPPLTLILLYLSTIYTR
WKVDVAGLLQCPILLVLTWADFLTLILPLTELKLYLTKVTRTDIERMGLG
IDYTRVDSIYVDSBEGVYLFPSROKAQGNFSILLPLIKATLISCVSMMQLIYKSY
LTDPMYMRKVIIEISGNTIISRLVAALIELMWSMEESKGLKFFYLLSRLRN
LIIHKHVRNEVAVSWGEEVYGMPIKNTIIKASTLSKSHCICTVCEGRWKGTC
PCGHRGPTCCGMSLADFEERHKRIFIRGNEFGMCSKQGRHRRFERMDRPSHRY
YCAENRHPABEGDFNAESMLGKITYPALMDGKVYDITENAGQVGSIPDTHAR
PCHILSGRMPFROEYGVQYTGQLFRLNLPVLTAKVLMVNLGMLNGLNIEGNLHL
GWILRGPAVCKKITHEKCHINIDKLTAFGLIMPRTTPRAPVPTSLLRVRLGLE
TGMATYHOGGISVDHVTAGKDLLVCDMSGRTRVVCOSNNRLTDEYEVKTDGCPD
GASSYVGRVGRKNEESKPTKMSGQVTSKNTADLTEMVKKITSNRNGDQKITLA
TCAGKTTLPKAVIEEIGRHKRVLLVPLRAAASVVOYRMLKHPISFNRLGDKME
GDMACTITPASYGYFCQPKRAAMVYSYIFDEYHCAATPEQLAIICKIHRFSES
IRVAMTATPAGSVTTTGKQKPIIEFTAPEYMGEDLGSQFLDIAGLKIPLVDMKNMT
LVEYTRNMAVEVAKKLKKNYSYISGDEPANLRVTSQSPYIVATNAIESGVN
LPDLTDVDTGLKCEKRVSVSSKPIFFVTGLKRAMVTVGQAGRGRVGRVGRYTR
SOETATGSKYHDLQQRIGIEDGINVTKSFREMYDWSLEYEEDSLTIQLEINLN
LLISELPAVKYNNIMARTDHPETQLAYNSYEQVPLFPKIRNGEYDLYNYSNLN
ARKLEDYPVYIYATDEDEDLAVDLGLDWDPDGQVQVETGKALKQVTLGSSAENALL

VALFGVYGVYGOALSKRRVPMITDIYITIEDORLEDTIHLQYAPNAIKTGDGTETELKELAS
GDVEKINGALSDYVAAGLEFVKSOAEKIKTAPLAKFENAEAGAGVQKFDISLTENKEE
IIRYGLWGTHTALYKSLAARLGHETAFATLVLKWAFGESVSUHVQKQALVDLVVYV
MKNPSPGDSSETQOEGRRVVASLFI SALATYTYKTYNHNLSKVLPALAYLPYATSA
LKMFPTPRLSVAIVSTTKTYKTLRSIRKKSODGLGTGISAANEIISQNPVSVGISVM
LGVFTAAHNAIENSSQRTLLMKVFKNFLDQAATDELVKENPEKILMALFAVQTI
GNPLRIYHLUYGVYKWEAKESRTAGNRLFTLIMPEAFELGLMDSGKIRNLSGN
YIIDLIIYGLHQNRLGKMKVLGAPAFSCDTPSDRIRLPTDNLVRYETPCOGY
EMAFKNVGGKLVKVESGPFKRNPRGPNVRYTKYDDNLRKIPKVKLEGQVE
HYKGVKTADYKSGKMLATDKVEHGVITLAKRYTGVGPNGLGDEONHRAV
ERCATITKNTVQFLKMKKCAFTYDLTISNLRLELHVRNMLEKEIPTALITWVL
AYTFVNDVGTIKPVLIGERVIPDPVDINDLOPEVQDTSVGTITIGRETLMITGTP
VLEKVEPDSADNSKVIKDEGNPGPGIQTHTLTEEIHNRDARFIFILGSRNSIS
NRAKTARNILNYTGNDRPRETRDMAARMLVALRDVPELSMDFKGTFLDREALE
ALSLOGPKQVOTKEAVRNLEOKKDVEIPNPFASDDPVFEVALANDKXYLVLDGVE
VKDQAKALGATDQTRILKEVGSRTYAMKLSWFLOASNKOMSLTLPFELLRCPPAT
KSNKGHMASAYQLAOGNNEPLGCGVHLGTIPARRVKIHPYEAULKLDHFEEEKAPR
VKDTVIREHNKMLIKTFOGNEPLKMLNPGKLSLOLDREGRKNIYHNOITMSS
AGIRLEKLPTRAQDTKTPEATIRDKIDKSENRONPELHNKLELFIHTIAQPTLKH
YGEVTFQLEAGINRKGAGLEKKNIGEVLDSEKHLVQLVRLDKAGRKIKYETAI
PKNEKRDVSDMOAGDLVVEKPRVIOYPEAKTRLAITKVYNNVQOOPVJPCYEGK
TPLFNIPIFKYRKEDWSENVAVSFDTKAWDTQVTSKDLQELICEIOKYVVKEMHFI
DTIDHTEVPVITADGEVIRNCGSGOPDTSAGNSMLNVLTMWAFCESTGVYK
SFNRVARIHVGCGDDFLITBKGGLAFANKGMQILHEAGAPKITEGEMKAVAFRED
IEFCSHTPVPVRNSDNTSSHMAGRDATVILSKMATRLDSSGERGTTAYEKAVAFSL
MYSNPLVRRICLIVLSQOPEPDSKHATYYKGDPIGAYKDVIGIRNLSKLRGTPEK
LANLNLISLTGLITWTKTSKRIIODCVAIKERGNMLVNADRLISKTLGHLIPDKGF
TLOGKHVYEQLOLQETNPVNGVCTERYKLGPIVNLRLRLKILLMTAVGVSS"
12509..12734

3'UTR 4030 a 2608 c 3293 g 2802 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 2,55e-63 Length: 12734
Score: 910.50 Matches: 178
Percent Similarity: 93.85% Conservative: 5
Best Local Similarity: 91.28% Mismatches: 9
Query Match: 88.23% Indels: 1
DB: 14 Gaps: 3

US-09-965-594-14 (1-197) x AF268278 (1-12734)
Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
Db 413 GGTAGTGTGTATGTGTAGATAATGTTTATCTGTAGTGTAGTATCACGGCGTCG 472
Qy 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41
Db 473 GCCCAGCAGCAGGAGGCGCTCTAGGGTGTAAAGATCACCACTGTGCTGGCGGGACAA 532
Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAlaThr 61
Db 533 AACCAAGTGGAGGTGAGGTCCAGATCGTGCACTGCTACCCAAACCTTCTTGGCAACG 592
Qy 62 CysIleAsnGlyValCysTrpThrValTyHHisGlyAlaGlyThrArgThrIleAlaSer 81
Db 593 TGCATCAATGGGGTATGCTGGACTGTCTACACCGGGCGGACGAGGACCATCGCATCA 652
Qy 82 ProLysGlyProValIleGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpPro 101
Db 653 CCCAAGGCTCTGTGATCCAGATGTATACCAATGTGGACCAAGACCTTGTGGGTGCCCC 712
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
Db 713 GCTCCTCAAGGTTCCTCGCTCATTCGACACCTGACCTGCGGCTCTCGGACCTTTACCTG 772
Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
Db 773 GTTACGAGCCAGCAACGCTATTCCTCCGCGCGCGGCGAGGTGATAGCAGGGGTAGCCTG 832
Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
Db 833 CTTTGGCCCCGGCCATTTCCTACATAAAGGCTCCTCTGGGGGTCCGCTGTGTGTGCCCC 892

Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181
Db 893 GCGGACACCGCGTGGGCTATTTCAGGCGCGGGTGTGCACCGCTGGAGTGCCCAAGCGG 952
Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196
Db 953 GTGGACTTATCCCTGTGAGAACCTAGAGACACACGACGATCC 997
RESULT 3
ARL18686
LOCUS ARL18686 5360 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 53 from patent US 6150087.
ACCESSION ARL18686
VERSION ARL18686.1 GI:14100596
KEYWORDS
SOURCE Unknown.
ORGANISM Unclonable.
REFERENCE 1 (bases 1 to 5360)
AUTHORS Chien,D.Y.
TITLE NANBV diagnostics and vaccines
JOURNAL Patent: US 6150087-A 53 21-NOV-2000;
FEATURES Location/Qualifiers
source 1..5360
BASE COUNT 1060 a 1623 c 1532 g 1145 t
ORIGIN
Alignment Scores:
Pred. No.: 5,32e-63 Length: 5360
Score: 901.50 Matches: 175
Percent Similarity: 90.20% Conservative: 9
Best Local Similarity: 85.78% Mismatches: 11
Query Match: 87.35% Indels: 9
DB: 6 Gaps: 1
US-09-965-594-14 (1-197) x ARL18686 (1-5360)
Qy 3 LysLysGlySerValValIleValGlyArgIleAsn----- 14
Db 867 CGCAGGGCGGGAGATCTGCTCGGGCCAGCGGATGGTAATGCTTCCAAAGGGGTGGAGG 926
Qy 15 ---LeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlu 33
Db 927 TTGCTGCGCCCATCATCAGGGGTACGCCAGCAGCAGACAGAGGGGCTCTCAGGTGCATAATC 986
Qy 34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 53
Db 987 ACCAGGCTTAATCGCGGGGACAAAAACCAAGTGGAGGGTGGAGTCCAGATTGTGCAACT 1046
Qy 54 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGly 73
Db 1047 GCTGCCCAACCTTCTCGCAACGTCATCAATGGGTGTGTGCTGTCTTACCAAGGG 1106
Qy 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 93
Db 1107 GCCGGAACGAGGACCATCGCTCACCAGGGTCTCATCCAGATGTATACCAATGTA 1166
Qy 94 AspLysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThr 113
Db 1167 GACCAAGACCTTCTGGGCTGGCCCGCTCCGCAAGGTAGCGCTCATTTGACACCTGCACT 1226
Qy 114 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 133
Db 1227 TGGCGCTCTCTCGGACCTTTACTTGGTCAGAGCAGCCGATGTCATCCCGGTGGCGGG 1286
Qy 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 153
Db 1287 CGGGGTGATAGCAGGGGACCGCTGCTGCGCCCGGCCCATTTCTTACTTGAAGGCTCC 1346
Qy 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaLysAla 173
Db 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaLysAla 173

```
Db      1347 TCGGGGGTCCGCTGTTGGCCCCGGGGCAGCGCTGGGCATATTTAGGGCCGCGTG 1406
Qy      174 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 193
Db      1407 TGCACCGGTGGAGTGGCTAAGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACC 1466
Qy      194 MetArgSerPro 197
Db      1467 ATGAGGTCCCG 1478

RESULT 4
LOCUS      106434
DEFINITION Sequence 48 from Patent EP 0318216.
ACCESSION 106434
VERSION 106434.1 GI:590311
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 5360)
AUTHORS Houghton,M., Choo,Q.-L. and Kuo,G.
TITLE Nanbv diagnostics and vaccines
JOURNAL Patent: EP 0318216-A1 48 31-MAY-1989;
FEATURES
    Location/Qualifiers
        source
            1..5360
                /organism="unknown"
BASE COUNT 1061 a 1623 c 1533 g 1143 t
ORIGIN

Alignment Scores:
Pred. No.: 5,32e-63 Length: 5360
Score: 901.50 Matches: 175
Percent Similarity: 90.20% Conservative: 9
Best Local Similarity: 85.78% Mismatches: 11
Query Match: 87.35% Indels: 9
DB: Gaps: 1

US-09-965-594-14 (1-197) x 106434 (1-5360)
Qy      3 LysLysGlySerValValIleValGlyArgIleAsn----- 14
Db      867 CGCAGGGGCGGGAGATGCTCGGGCCAGCGCATGGAAATGCTCTCAAGGGGTGGAGG 926
Qy      15 ---LeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGlyCysGlnGlu 33
Db      927 TTGCTGGCGCCCATCAGCGGTACGCCAGCAGACAAGGGGCTCCTAGGTGCATAATC 986
Qy      34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 53
Db      987 ACCAGCCTAACTGGCGGGACAAAACCAAGTGGAGGGTGAGTCCAGATTGTGTCAACT 1046
Qy      54 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGly 73
Db      1047 GCTGCCCAAACTTCTGGCAACTGCATCAATATGGGGTGTCTGGACTGTCTACCAAGGG 1106
Qy      74 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 93
Db      1107 GCCGGAACGAGGACCATCGCGTCACCAAGGTCTGTATCCAGATGTATACCAATGTA 1166
Qy      94 AspLysAspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThr 113
Db      1167 GACCAAGACCTTGTGGGCTGGCCGCTCCGCAAGTAGCGCTCATTTGACACCTGCAC 1226
Qy      114 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 133
Db      1227 TCGGGCTCTTCGAGACCTTTACTGTGTACGAGGACGCCGATGTCAATTCCTGGCCCGG 1286
Qy      134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 153
Db      1287 CGGGGTGATAGCAGGGGACGCTGCTGTCGCCCGGCCCATTTCTTACTTTGAAGGCTCC 1346
Qy      154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 173
```

```
Db      1347 TCGGGGGTCCGCTGTTGGCCCCGGGGCAGCGCTGGGCATATTTAGGGCCGCGTG 1406
Qy      174 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 193
Db      1407 TGCACCGGTGGAGTGGCTAAGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACC 1466
Qy      194 MetArgSerPro 197
Db      1467 ATGAGGTCCCG 1478

RESULT 5
LOCUS      109328
DEFINITION Sequence 8 from Patent WO 8904669.
ACCESSION 109328
VERSION 109328.1 GI:587963
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 5360)
AUTHORS Houghton,M., Choo,Q.-K. and Kuo,G.
JOURNAL Patent: WO 8904669-A 8 01-JUN-1989;
FEATURES
    Location/Qualifiers
        source
            1..5360
                /organism="unknown"
BASE COUNT 1061 a 1623 c 1533 g 1143 t
ORIGIN

Alignment Scores:
Pred. No.: 5,32e-63 Length: 5360
Score: 901.50 Matches: 175
Percent Similarity: 90.20% Conservative: 9
Best Local Similarity: 85.78% Mismatches: 11
Query Match: 87.35% Indels: 9
DB: Gaps: 1

US-09-965-594-14 (1-197) x 109328 (1-5360)
Qy      3 LysLysGlySerValValIleValGlyArgIleAsn----- 14
Db      867 CGCAGGGGCGGGAGATGCTCGGGCCAGCGCATGGAAATGCTCTCAAGGGGTGGAGG 926
Qy      15 ---LeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGlyCysGlnGlu 33
Db      927 TTGCTGGCGCCCATCAGCGGTACGCCAGCAGACAAGGGGCTCCTAGGTGCATAATC 986
Qy      34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 53
Db      987 ACCAGCCTAACTGGCGGGACAAAACCAAGTGGAGGGTGAGTCCAGATTGTGTCAACT 1046
Qy      54 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGly 73
Db      1047 GCTGCCCAAACTTCTGGCAACTGCATCAATATGGGGTGTCTGGACTGTCTACCAAGGG 1106
Qy      74 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 93
Db      1107 GCCGGAACGAGGACCATCGCGTCACCAAGGTCTGTATCCAGATGTATACCAATGTA 1166
Qy      94 AspLysAspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThr 113
Db      1167 GACCAAGACCTTGTGGGCTGGCCGCTCCGCAAGTAGCGCTCATTTGACACCTGCAC 1226
Qy      114 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 133
Db      1227 TCGGGCTCTTCGAGACCTTTACTGTGTACGAGGACGCCGATGTCAATTCCTGGCCCGG 1286
Qy      134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 153
Db      1287 CGGGGTGATAGCAGGGGACGCTGCTGTCGCCCGGCCCATTTCTTACTTTGAAGGCTCC 1346
Qy      154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 173
```

```
|||||
Db 1347 TCGGGGGTCCGCTGTGTGTCGCCGGGACGCCGTGGCATATTTAGGCGCGGGTG 1406
Qy 174 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 193
Db 1407 TGCACCGGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGAGAACCTTAGACAACC 1466
Qy 194 MetArgSerPro 197
Db 1467 ATGAGGTCCCGC 1478

RESULT 6
LOCUS AR118692 6785 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 65 from patent US 6150087.
ACCESSION AR118692
VERSION AR118692.1 GI:14100602
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 6785)
AUTHORS Chien,D.Y.
TITLE NAMV diagnostics and vaccines
JOURNAL Patent: US 6150087-A 65 21-NOV-2000;
FEATURES
    Location/Qualifiers
        1..6785
            /organism="unknown"
BASE COUNT 1392 a 2050 c 1914 g 1429 t
ORIGIN

Alignment Scores:
Pred. No.: 6.86e-63 Length: 6785
Score: 901.50 Matches: 175
Percent Similarity: 90.20% Conservative: 9
Best Local Similarity: 85.78% Mismatches: 11
Query Match: 87.35% Indels: 9
DB: 6 Gaps: 1

US-09-965-594-14 (1-197) x AR118692 (1-6785)
Qy 3 LysLysGlySerValValIleValGlyArgIleAsn----- 14
Db 1140 CGCAGGCGCGGGAGATCTCTCGGCCAGCCGCGAATGTCTCCAAAGGGTGGAGG 1199
Qy 15 ---LeuSerGlyAspThrAlaIleGlnThrArgGlyGluGlyCysGlnGlu 33
Db 1200 TTGCTGGGCCCATCATCGGCGTAGCCCGCAGACAGAGGGGCTCCTTAGGGTGCATAATC 1259
Qy 34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 53
Db 1260 ACCAGCCTAATCGCGCGGACAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGCAACT 1319
Qy 54 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValThrHisGly 73
Db 1320 GCTGCCCAAACTTCTCGGCAACGTGCATCAATGGGGTGTGCTGACTGTCTACACGGG 1379
Qy 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 93
Db 1380 GCCGGAACGAGGACCATCGCGCTACCCCAAGGGTCTGTCTATCCAGATGTATACCAATGTA 1439
Qy 94 AspLysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThr 113
Db 1440 GACCAAGACCTTGTGGGTGGCCGCTCCGCAAGGTAGCCGCTCATTTGACACCCCTGCACT 1499
Qy 114 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 133
Db 1500 TCGGGCTCTCGGACCTTTACCTGGTCAGGAGGACGCCGATGTTCATTCCTCGCGCGG 1559
Qy 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 153
Db 1560 CGGGGTGATACAGGGGACGCGCTGCTGTGCGCCCGGCCCATTTCTCTACTTGAAGGCTCC 1619
```

```
Qy 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaVal 173
Db 1620 TCGGGGGTCCGCTGTGTGTCGCCGGGACGCCGTGGCATATTTAGGCGCGGGTG 1679
Qy 174 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 193
Db 1680 TGCACCGGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGAGAACCTTAGACAACC 1739
Qy 194 MetArgSerPro 197
Db 1740 ATGAGGTCCCGC 1751

RESULT 7
LOCUS I06440 6785 bp DNA linear PAT 02-DEC-1994
DEFINITION Sequence 54 from Patent EP 0318216.
ACCESSION I06440
VERSION I06440.1 GI:590312
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 6785)
AUTHORS Houghton,M., Choo,Q.-L. and Kuo,G.
TITLE Nanbv diagnostics and vaccines
JOURNAL Patent: EP 0318216-A1 54 31-MAY-1989;
FEATURES
    Location/Qualifiers
        1..6785
            /organism="unknown"
BASE COUNT 1392 a 2050 c 1914 g 1429 t
ORIGIN

Alignment Scores:
Pred. No.: 6.86e-63 Length: 6785
Score: 901.50 Matches: 175
Percent Similarity: 90.20% Conservative: 9
Best Local Similarity: 85.78% Mismatches: 11
Query Match: 87.35% Indels: 9
DB: 6 Gaps: 1

US-09-965-594-14 (1-197) x I06440 (1-6785)
Qy 3 LysLysGlySerValValIleValGlyArgIleAsn----- 14
Db 1140 CGCAGGCGCGGGAGATCTCTCGGCCAGCCGCGAATGTCTCCAAAGGGTGGAGG 1199
Qy 15 ---LeuSerGlyAspThrAlaIleGlnThrArgGlyGluGlyCysGlnGlu 33
Db 1200 TTGCTGGGCCCATCATCGGCGTAGCCCGCAGACAGAGGGGCTCCTTAGGGTGCATAATC 1259
Qy 34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 53
Db 1260 ACCAGCCTAATCGCGCGGACAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGCAACT 1319
Qy 54 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValThrHisGly 73
Db 1320 GCTGCCCAAACTTCTCGGCAACGTGCATCAATGGGGTGTGCTGACTGTCTACACGGG 1379
Qy 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 93
Db 1380 GCCGGAACGAGGACCATCGCGCTACCCCAAGGGTCTGTCTATCCAGATGTATACCAATGTA 1439
Qy 94 AspLysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThr 113
Db 1440 GACCAAGACCTTGTGGGTGGCCGCTCCGCAAGGTAGCCGCTCATTTGACACCCCTGCACT 1499
Qy 114 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 133
Db 1500 TCGGGCTCTCGGACCTTTACCTGGTCAGGAGGACGCCGATGTTCATTCCTCGCGCGG 1559
Qy 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 153
Db 1560 CGGGGTGATACAGGGGACGCGCTGCTGTGCGCCCGGCCCATTTCTCTACTTGAAGGCTCC 1619
```

QY 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaVal 173
Db 1620 TCGGGGGTCCGCTGTGTGCCCCGGGGACGGCGTGGCATATTTAGGGCCGGGTG 1679
QY 174 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 193
Db 1680 TGCACCGGTGGAGTGGCTAAGGGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACC 1739
QY 194 MetArgSerPro 197
Db 1740 ATGAGTCCCG 1751
RESULT 8
LOCUS I09329 6785 bp DNA linear PAT 02-DEC-1994
DEFINITION Sequence 10 from Patent WO 8904669.
ACCESSION I09329
VERSION I09329.1 GI:587964
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 6785)
AUTHORS Houghton,M., Choo,Q.-K. and Kuo,G.
JOURNAL Patent: WO 8904669-A 10 01-JUN-1989;
FEATURES
Location/Qualifiers
1..6785
/organism="unknown"
BASE COUNT 1392 a 2050 c 1914 g 1429 t
ORIGIN
Alignment Scores:
Pred. No.: 6.86e-63 Length: 6785
Score: 901.50 Matches: 175
Percent Similarity: 90.20% Conservative: 9
Best Local Similarity: 85.78% Mismatches: 11
Query Match: 87.35% Indels: 9
DB: Gaps: 1
US-09-965-594-14 (1-197) x I09329 (1-6785)
QY 3 LysLysGlySerValValIleValGlyArgIleAsn----- 14
Db 1140 CGCAGGGCCGGGAGATGCTCGGGCCACCGCATGGATGCTCCAGGGGTGGAGG 1199
QY 15 ---LeuSerGlyAspThrAlaValAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlu 33
Db 1200 TTGCTGGGGCCATCAGCGGCTAGCCCGACAGACAGGGGCTCTAGGGTGCATAATC 1259
QY 34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 53
Db 1260 ACCAGCTTAATCGCCGGGACAAAACCAAGTGGGGTGGAGTGGAGTGTGTCAACT 1319
QY 54 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyHisGly 73
Db 1320 GCTGCCAACCCTTCCTGGCAACGTGATCAATGGGGTGTGCTGGACTGTCTACCGGG 1379
QY 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyThrAsnVal 93
Db 1380 GCCGGAACGAGGACCATCGGCTCACCCAAAGGTCCTGTCATCCAGATGTATACCAATGTA 1439
QY 94 AspLysAspLeuValGlyTTPProAlaProGlnGlySerArgSerLeuThrProCysThr 113
Db 1440 GACCAAGACCTTGTGGGCTGGCCGGCTCCGCAAGGTAGCCGCTCATTTGACACCCCTGCACT 1499
QY 114 CysGlySerSerAspLeuValThrArgHisAlaAspValIleProValArgArg 133
Db 1500 TGGCGCTCCTCGGACCTTACCTGGTCACGAGCAGCCGATGTCTATCCCGTGGCCGG 1559
QY 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyLeuLysGlySer 153
Db 1560 CGGGGTGATAGAGGGGAGCGCTGCTCGCCCGGGCCCATTTCTTCTACTTGAAGGCTCC 1619

QY 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaVal 173
Db 1620 TCGGGGGTCCGCTGTGTGCCCCGGGGACGGCGTGGCATATTTAGGGCCGGGTG 1679
QY 174 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 193
Db 1680 TGCACCGGTGGAGTGGCTAAGGGGTGGACTTTATCCCTGTGGAGACAACC 1739
QY 194 MetArgSerPro 197
Db 1740 ATGAGTCCCG 1751
RESULT 9
LOCUS AR118696 7310 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 74 from patent US 6150087.
ACCESSION AR118696
VERSION AR118696.1 GI:14100606
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 7310)
AUTHORS Chien,D.Y.
JOURNAL NABV diagnostics and vaccines
PATENT: US 6150087-A 74 21-NOV-2000;
FEATURES
Location/Qualifiers
1..7310
/organism="unknown"
BASE COUNT 1495 a 2220 c 2056 g 1539 t
ORIGIN
Alignment Scores:
Pred. No.: 7.44e-63 Length: 7310
Score: 901.50 Matches: 175
Percent Similarity: 90.20% Conservative: 9
Best Local Similarity: 85.78% Mismatches: 11
Query Match: 87.35% Indels: 9
DB: Gaps: 1
US-09-965-594-14 (1-197) x AR118696 (1-7310)
QY 3 LysLysGlySerValValIleValGlyArgIleAsn----- 14
Db 1665 CGCAGGGCCGGGAGATGCTCGGGCCACCGCATGGATGCTCCAGGGGTGGAGG 1724
QY 15 ---LeuSerGlyAspThrAlaValAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlu 33
Db 1725 TTGCTGGGGCCATCAGCGGCTAGCCCGACAGACAGGGGCTCTAGGGTGCATAATC 1784
QY 34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 53
Db 1785 ACCAGCTTAATCGCCGGGACAAAACCAAGTGGGGTGGAGTGGAGTGTGTCAACT 1844
QY 54 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyHisGly 73
Db 1845 GCTGCCAACCCTTCCTGGCAACGTGATCAATGGGGTGTGCTGGACTGTCTACCGGG 1904
QY 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyThrAsnVal 93
Db 1905 GCCGGAACGAGGACCATCGGCTCACCCAAAGGTCCTGTCTATCCAGATGTATACCAATGTA 1964
QY 94 AspLysAspLeuValGlyTTPProAlaProGlnGlySerArgSerLeuThrProCysThr 113
Db 1965 GACCAAGACCTTGTGGGCTGGCCGGCTCCGCAAGGTAGCCGCTCATTTGACACCCCTGCACT 2024
QY 114 CysGlySerSerAspLeuValThrArgHisAlaAspValIleProValArgArg 133
Db 2025 TCGGCTCCTCGGACCTTACCTGGTCACGAGGACGCGGATGTCTATCCCGTGGCCGG 2084
QY 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyLeuLysGlySer 153


```
Db 2085 CGGGGTGATAGCAGGGCAGCGTCTGCTCGCCCGGCCCATTTTCTACTTGAAGGCTCC 2144
Qy 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaVal 173
Db 2145 TCGGGGGTCCGCTGTGTGCCCCGGGGCAGCGCTGGGCATATTATTAGGCCGCGGTG 2204
Qy 174 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThr 193
Db 2205 TGCACCGGTGAGTGGCTAAGCGGTGACTTTATCCCTCTGGAGAACCTAGACACAAC 2264
Qy 194 MetArgSerPro 197
Db 2265 ATGAGTCCCCG 2276

RESULT 10
LOCUS I09331 7310 bp DNA linear PAT 02-DEC-1994
DEFINITION Sequence 15 from Patent WO 8904669.
ACCESSION I09331
VERSION I09331.1 GI:587966
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 7310)
Houghton, M., Choo, Q.-K. and Kuo, G.
JOURNAL Patent: WO 8904669-A 15 01-JUN-1989;
FEATURES
Location/Qualifiers
source 1..7310
/organism="unknown"

BASE COUNT 1495 a 2218 c 2058 g 1539 t
ORIGIN

Alignment Scores:
Pred. No.: 7,44e-63 Length: 7310
Score: 901.50 Matches: 175
Percent Similarity: 90.20% Conservative: 9
Best Local Similarity: 85.78% Mismatches: 11
Query Match: 87.35% Indels: 9
DB: 6 Gaps: 1

US-09-965-594-14 (1-197) x I09331 (1-7310)

Qy 3 LysLysGlySerValValIleValGlyArgIleAsn----- 14
Db 1665 CGCAGGGCGGGAGATACTGTCGCGCCAGCCGATGGAATGCTCTCAAGGGGTGGAGG 1724
Qy 15 ---LeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlu 33
Db 1725 TTGCTGGCCCGCCATCAGCGCGTACCGCCAGCAGACAGGAGGCTCCTAGGTGCTAATC 1784
Qy 34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 53
Db 1785 ACCAGCCTAACTGCGCGGCACAAAACCAAGTGAGGCTGAGGTCCAGATTGTGCAACT 1844
Qy 54 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValThHisGly 73
Db 1845 GTGTGCCCAACCTCTCGGCAACGTGCATCAATGGGGGTGCTGCTGACGTCTACACCGG 1904
Qy 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 93
Db 1905 GCGGACGAGGACCATCGCTCACCAGGGTCTGTCATCCAGATGTATACCANTGTA 1964
Qy 94 AspLysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThr 113
Db 1965 GACCAAGACCTTGTGGCTGGCCGCTCGCAAGGTAGCGCTCATTGACACCGCTGCAC 2024
Qy 114 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArg 133
Db 2025 TCGCGTCTCGACCTTTACCTGCTGACGAGCGACCGCATGCTCATTCCTCCGTCGCGG 2084
Qy 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 153
```

```
Db 2085 CGGGGTGATAGCAGGGCAGCGTCTGCTCGCCCGGCCCATTTTCTACTTGAAGGCTCC 2144
Qy 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaVal 173
Db 2145 TCGGGGGTCCGCTGTGTGCCCCGGGGCAGCGCTGGGCATATTATTAGGCCGCGGTG 2204
Qy 174 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThr 193
Db 2205 TGCACCGGTGAGTGGCTAAGCGGTGACTTTATCCCTGTGGACAACCTAGACACAAC 2264
Qy 194 MetArgSerPro 197
Db 2265 ATGAGTCCCCG 2276

RESULT 11
LOCUS HPCPOLYP 7310 bp ss-RNA linear VRL 02-AUG-1993
DEFINITION Hepatitis C virus polyprotein gene, partial cds.
ACCESSION M32084
VERSION M32084.1 GI:329875
KEYWORDS polyprotein.
SOURCE Hepatitis C virus
ORGANISM Hepatitis C virus
REFERENCE
1 (bases 1 to 7310)
Choo, Q.-L., Richman, K. and Han, J.
JOURNAL The nucleotide sequence of the Hepatitis C viral genome
Unpublished (1990)
COMMENT Original source text: Hepatitis C virus, cDNA to viral RNA, clones K9-1 through 15e, isolated from chimpanzee (individual 910) blood plasma.
Draft entry and printed sequence for [1] kindly submitted by M. Houghton, 22-FEB-1990. Chiron Corporation, 4560 Horton Street, Emeryville CA 94608.
FEATURES
Location/Qualifiers
1..7310
/organism="Hepatitis C virus"
/mol_type="genomic RNA"
/db_xref="taxon:11103"
<1..>7310
/notes="polyprotein"
/codon_start=3
/protein_id="AAA45677.1"
/db_xref="GI:329876"
/translation="GCPEERLASRLPTDFDQMGPISYANGSGPQDPQPCWYHPKPKC
GIVPAKSCGVPVCFPTSPVYGTDRSGAPTSWGENDTDFVLNNTRPPLGNWPGC
TWNNSGFTKVCAGPCPVIGGAGNNTLHCPTDCFRKHPDATYSGSGSPMITPRCLVD
YPRLHNPCTINYTIETKIRMYGVGHELEACNKTGERCDLDRBSELSPLILT
TTQWQVPCSTFTLPALSTGLHLHQNIVDVQILYGVSSIASNAIKWIFVYLLFLL
ADARVCSLMMMLLIISOAAALENIVLNAASLAGTHGLVSLVFFCFAMYLUKRWVP
GAYTTTGYMPLLLLLLALPORAYALDTEVAASCGGVLVGLMALTLSPPYKRYISWC
LMWQYLFVRVEAQLHWPPLNVRGRDAVILMCVHPTLVFDITKLLAVFGPLW
ILQASLLKPYFVRVQGLLRFCALARKMTGGHVQMVIIKLGALGTYYVYHLTPRD
WANGRLDLAVAVEPVVSOMETKLITMGADTAACGDIINGLPVARRGRETLLGPAD
GMYSKWRLLAPITAYAQTRGLLGLIITSLGRDNKNQVEGVQIVYSTAAQPLATCI
NGVCVTYRGAGTRTASPRGSPVQMTINVDQLVGWPAPQSGRSPTCTCGSSDLX
VTRHADVIPVRRRGDSGLSPRISLYLKSGSGGLCPAGHAGVIFRAAYCTGSLA
KAYVLNLSVAATLFGFAMYSKAHGDIPNIRGTVRTITTSPTITTYGKFLDAAYAQ
YKVDLPENLETTMRSPVTONSSPVVPSQFVAHLHAPTHAPTSKSTKPAAYAAQG
GGAYDIIICDECHSDATSLIGTLVDOAETAGARLVATATATPGSVVPHNTEE
VALSTGEIPPYGKAIPLEVIKGRHLIFCHSKKCDLAKLVALGINAVAYRGLD
VSVPTSGDVVATDALMTGTGDFSDVIDCNTCTVDFSLQPTFTTETILPQD
AVSKTRRGRTGKPGIYFRVAPRPSGMSDFSVLCYCDAGCAWELTAPETVR
LRAYMTPLGPGVQDHLFEWEGVTGLTHDAHFTSOTKQSGENLPYLVAYQATCAR
AQAPPSWDQMWKCLIRLAPTLHGPTPLLYRLGAVONEITLTHPTVKYIMTCSADLE
VVTSTVYGGVLAALAAVCLSTGCVIVGVVLSGKPAIIPDREVLVREFDEBES
QHLPSTYEQMMLAEQFKQALGLQIASROAEVIAFVQTNWQKLETFWAKHMRNFIS
GIQYLAGSLTLPGNPAISLMFAETAVTSPITTSQTLNLLGGWVAQALAAPGAATA
FVAGLAGLAAIGSVGLGKVLIDILAGYAGAGALVAFKIMSGEVPSTEDLVNLLPAI
LSPGALVGVVCAAILRRHVGPAGAVQNMNRLIAFASRGNVSPHYVPESDAAARV
TAILSLITVOLLRLRHVLSSECTPCSGSLRDLTDWATCEVSLDFKTLAKLWPO
LPGLPFPVSCORGKGVWRVDGIMHTRCHGAEITGVKNGKTRIVGPRICRNWSGTF
```

PINATYTGCTPLPAPNTFALWRVSAEYVEIROVGFHYVTGTTDNLKPQCVPS
PEFFTELDGRLRHPAPCKPLREEVFRVGLHEYPVGSQLPCEPEPDVAULTSMLT
DPSHITAAGRLARGLSPVASSASQLSAPSKATCTANHDSQPDALTEANLLWR
QMGNGNITRVSENGKVVILDSFDPLVAREDEREISVPAEILRKSRFPQALPVWARP
YNPPLVETWKKPDYEPVPHGCPPLPPKSPVPPPKRTTVPVLESTLSTALAEATR
SFGSSSTGIGDNTTSSPAPSCPDSDAESYSSMPLEGEPPDLSDSGWSVT
SSEANEDVCCSMYSWTGALVTPCAAEQKLPIINALSNLLRHNLVYITTSRSAC
QKRVTFDRQLVDSHYODVLKEYKAASRKANLSEVEACSLTPPHSAKSFYGG
AKDVRCHARKMALYINSVWKOLLNDNVTPIIDTINAKNEVFCVQPEKGRKPARLIV
POLGVRCCKMAYLVDVTKPLAVMGSSYGFQYSPQORVFLVOAKWSKKTPMGFSDY
TRCFDSTVEDIRTEBALYOCDDLDQARVAIKSLTERLYVGGPLTNSRNGCYRR
CRAGVLTSQNTLITIKARAACRAGLQCTMLVCGDDLVICESAGVEDAASL
RAFTAMRTYSAPPDGPPEYDELEITSCSNVSVAHGAGKRVYIITRDTPLAR
AAMETARHPVNTWGNLIMEFAPTLWRMILMTHFFSVLIARDQLEQALDEIYGACY
STIEPLDLPPIIQR*

BASE COUNT 1495 a 2218 c 2058 g 1539 t
ORIGIN

Alignment Scores:
Pred. No.: 7,44e-63 Length: 7310
Score: 901.50 Matches: 175
Percent Similarity: 90.20% Conservative: 9
Best Local Similarity: 85.78% Mismatches: 11
Query Match: 87.35% Indels: 9
DB: 14 Gaps: 1

US-09-965-594-14 (1-197) x HPCPOLYP (1-7310)

QY 3 LysLysGlySerValValIleValGlyArgIleAsn----- 14
Db 1665 CGCAGGGCGGGAGATCTGCTCGGGCCAGCGGATGGTCTCCAAAGGGGTGGAGG 1724
QY 15 ---LeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGluCysGlnGlu 33
Db 1725 TTGCTGGCGCCCATCATCGGCGGTACGCCAGCAGACAAAGGGGCTCTCTAGGTGTCATAATC 1784
QY 34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 53
Db 1785 ACCAGCTAACTGGCGGGGACAAAACCAAGTGGAGGTGAGTGCAGATTGTGTCAACT 1844
QY 54 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGly 73
Db 1845 GCTGCCCAAACTTCTGTGCAACGTGCATCAATGGGTGTGCTGGACGTCTTACCACGGG 1904
QY 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 93
Db 1905 GCGGAACAGGAGCACCATCGCTCACCAAGGTCTCTGCATCCAGATGTATACCAATGTA 1964
QY 94 AspLysAspLeuValGlyTrpProAlaProGlnGlnSerArgSerLeuThrProCysThr 113
Db 1965 GACCAAGACCTTGTGGGCTGGCGGCTCCGCAAGGTAGCGCTCATTTGACACCTTGCAC 2024
QY 114 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 133
Db 2025 TGGGGCTCTCGGACCTTTACCTGTGTGTCAGGAGCAGCGCGATGTATTCCCGTGGCGGG 2084
QY 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 153
Db 2085 CGGGGTGATAGCAGGGGAGCGCTGTCTCGCGCGGCCCATTTCTTACTTTGAAGGCTCC 2144
QY 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 173
Db 2145 TCGGGGGGTCCGCTGTGTGCGCGGGGACCGGTGGGATATTTAGGCGCGGGT 2204
QY 174 CysThrArgGlyValAlaLysAlaValAspPheIleProValIleSerLeuGluThrThr 193
Db 2205 TGCACCGGTGAGTGGCTAAGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAAC 2264
QY 194 MetArgSerPro 197
Db 2265 ATGAGGTCCCGG 2276

RESULT 12

ARI18703
LOCUS ARI18703 8316 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 88 from patent US 6150087.
ACCESSION ARI18703
VERSION ARI18703.1 GI:14100613
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 8316)
AUTHORS Chien,D.Y.
TITLE NAMBV diagnostics and vaccines
JOURNAL Patent: US 6150087-A 88 21-NOV-2000;
FEATURES Location/Qualifiers
source 1..8316
BASE COUNT 1671 a 2529 c 2345 g 1771 t
ORIGIN

Alignment Scores:
Pred. No.: 8,55e-63 Length: 8316
Score: 901.50 Matches: 175
Percent Similarity: 90.20% Conservative: 9
Best Local Similarity: 85.78% Mismatches: 11
Query Match: 87.35% Indels: 9
DB: 6 Gaps: 1

US-09-965-594-14 (1-197) x ARI18703 (1-8316)

QY 3 LysLysGlySerValValIleValGlyArgIleAsn----- 14
Db 2671 CGCAGGGCGGGAGATCTGCTCGGGCCAGCGGATGGTGTCTCCAAAGGGGTGGAGG 2730
QY 15 ---LeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGluCysGlnGlu 33
Db 2731 TTGCTGGCGCCCATCATCGGCGGTACGCCAGCAGACAAAGGGGCTCTCTAGGTGTCATAATC 2790
QY 34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 53
Db 2791 ACCAGCTAACTGGCGGGGACAAAACCAAGTGGAGGTGAGTGTGTCAGATTTGTGTCAACT 2850
QY 54 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGly 73
Db 2851 GCTGCCCAAACTTCTGTGCAACGTGCATCAATGGGTGTGCTGGACTGTCTACACGGG 2910
QY 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 93
Db 2911 GCGGAACAGGAGCACCATCGCTCACCAAGGTCTCTGCATCCAGATGTATACCAATGTA 2970
QY 94 AspLysAspLeuValGlyTrpProAlaProGlnGlnSerArgSerLeuThrProCysThr 113
Db 2971 GACCAAGACCTTGTGGGCTGGCGGCTCCGCAAGGTAGCGCTCATTTGACACCTTGCAC 3030
QY 114 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 133
Db 3031 TGGGGCTCTCGGACCTTTACCTGTGTGTCAGGAGCAGCGCGATGTATCCCGTGGCGGG 3090
QY 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 153
Db 3091 CGGGGTGATAGCAGGGGAGCGCTGTCTCGCGCGGCCCATTTCTTACTTTGAAGGCTCC 3150
QY 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 173
Db 3151 TCGGGGGGTCCGCTGTGTGCGCGGGGACCGGTGGGATATTTAGGCGCGGGT 3210
QY 174 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 193
Db 3211 TGCACCGGTGAGTGGCTAAGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAAC 3270
QY 194 MetArgSerPro 197
Db 3271 ATGAGGTCCCGG 3282

ARL18728	ARL18728	8987 bp	DNA	linear	PAT 16-MAY-2001
LOCUS	Sequence 137 from patent US 6150087.				
DEFINITION	ARL18728				
ACCESSION	ARL18728				
VERSION	ARL18728.1	GI:14100638			
KEYWORDS	Unknown.				
SOURCE	Unknown.				
ORGANISM	Unclassified.				
REFERENCE	1 (bases 1 to 8987)				
AUTHORS	Chien, D.Y.				
TITLE	NANBV diagnostics and vaccines				
JOURNAL	Patent: US 6150087-A 137 21-NOV-2000;				
FEATURES	Location/Qualifiers				
source	1..8987				
	/organism="unknown"				
BASE COUNT	1807 a 2735 c 2547 g 1898 t				
ORIGIN					
Alignment Scores:					
Pred. No.:	9.3e-63	Length:	8987		
Score:	901.50	Matches:	175		
Percent Similarity:	90.20%	Conservative:	9		
Best Local Similarity:	85.78%	Mismatches:	11		
Query Match:	87.35%	Indels:	9		
DB:	6	Gaps:	1		
US-09-965-594-14 (1-197) x ARL18728 (1-8987)					
Qy	3 LysLysGlySerValValIleValGlyArgIleAsn-----				14
Db	3013 CGCAGGGCGCGGAGATGACTCTCGGGCCAGCCGATGGAAATGTCTCCAAAGGGGTGGAGG				3072
Qy	15 ---LeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlu				33
Db	3073 TTGCTGGCGCCCATCAGCGCGTACGCCACACAGAGGGCGCTCTCTAGGTGCTCATAATC				3132
Qy	34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr				53
Db	3133 ACCAGCCTACTTGGCGGGCAAAACCAAGTGGAGGTGAGGTCCAGATTGTGTCAACT				3192
Qy	54 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGly				73
Db	3193 GCTGCCCAACACTTCTGGCCACGTGCTCAATGGGGTGTGCTGGAGTGTCTACACCGGG				3252
Qy	74 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal				93
Db	3253 GCGGGAACGAGGACCATTCGGCTCACCAAGGGTCTGTCTATCCAGATGTATACCAATGA				3312
Qy	94 AspLysAspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThr				113
Db	3313 GACCAAGACCTTGTGGCTGGCCCGTCCGCAAGTAGCGCGCTATTGACACCTTGCACT				3372
Qy	114 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValIleArg				133
Db	3373 TGGCGCTCTCGGACCTTACCTGGTCACGAGGCACGCCGATGCTATCCGCTGGCGCGG				3432
Qy	134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer				153
Db	3433 CGGGGTGATACGAGGGCAGCTGCTCGCCCGCGGCCATTTCTACTTGAAGAGGCTCC				3492
Qy	154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal				173
Db	3493 TCGGGGGGTTCGCTGTTGTGCCCCCGGGGACGCCGTGGGCATATTATGGCGCGCGGTG				3552
Qy	174 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr				193
Db	3553 TGCACCGGTGGATGGCTAGGGCGGTGGACTTTATCTCTGTGGAGACACCTAGAGCAACC				3612
Qy	194 MetArgSerPro 197				
Db	3613 ATGAGGTCCCGG 3624				

```
Db      3932 ATGAGGTCCCG 3943
RESULT 15
AR118723/c
LOCUS   AR118723          9185 bp    DNA          linear    PAT 16-MAY-2001
DEFINITION   Sequence 123 from patent US 6150087.
ACCESSION   AR118723
VERSION     AR118723.1  GI:14100633
KEYWORDS    .
SOURCE      .
ORGANISM    .
REFERENCE   1 (bases 1 to 9185)
AUTHORS     Chien,D.Y.
TITLE       NANBV diagnostics and vaccines
JOURNAL     Patent: US 6150087-A 123 21-NOV-2000;
FEATURES    Location/Qualifiers
             source          1..9185
BASE COUNT  1938 a 2608 c 2790 g 1849 t
ORIGIN

Alignment Scores:
Pred. No.:          9,52e-63          Length:          9185
Score:             901.50             Matches:         175
Percent Similarity: 90.20%             Conservative:    9
Best Local Similarity: 85.78%           Mismatches:     11
Query Match:       87.35%             Indels:         9
DB:                6                  Gaps:          1

US-09-965-594-14 (1-197) x AR118723 (1-9185)
QY      3 LysLysGlySerValIleValGlyArgIleAsn----- 14
Db      5854 CGCAGGGCGGGAGATACTGCTCGGCCAGCGATGGAATGGTCTCCAAGGGGTGGAGG 5795
QY      15 ---LeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlu 33
Db      5794 TTGCTGGCGGCCATCAGCGCGGTACCGCCAGCAGACAGGGGCCCTCCTAGGGTGCATAATC 5735
QY      34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 53
Db      5734 ACCAGCCTAACTGGCGGGGACAAAACCAAGTGGAGGTGAGGTCCAGATTGTGTCAACT 5675
QY      54 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysIleThrValTyrHisGly 73
Db      5674 GCTGCCCAAACTCTCGCAACGTGCATCAATGGGTGTGCTGGACTGCTTACCACGGG 5615
QY      74 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 93
Db      5614 GCCGAAGCAGGAGCATCGCGTCACCCAAAGGTCTGTCTATCCAGATGATACCAATGTA 5555
QY      94 AspLysAspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThr 113
Db      5554 GACCAAGACCTTGTGGGTGGCGCCGCTCCGCAAGGTAGCCGCTCATTCACACCTGCACT 5495
QY      114 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 133
Db      5494 TCGGGCTCTCGGACCTTTACCTGTGTCAAGGACGCGCATGTCATTCCCGTGGCGCGG 5435
QY      134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 153
Db      5434 CGGGGTGATAGAGGGGACGCTGTGTGCGCCCGGGGACGCGGTGGGCATATTAGGGCCCGGTG 5375
QY      154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 173
Db      5374 TCGGGGGGTCCGCTGTGTGTCGCCCGGGGACGCGGTGGGCATATTAGGGCCCGGTG 5315
QY      174 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 193
Db      5314 TGCACCCGTGGAGTGGCTAAGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACC 5255
QY      194 MetArgSerPro 197
```

```
|||||
Db      5254 ATGAGGTCCCG 5243
Search completed: August 31, 2003, 00:46:04
Job time : 2584.57 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 30, 2003, 19:13:57 ; Search time 182.939 Seconds
(without alignments)
2906.924 Million cell updates/sec

Title: US-09-965-594-14

Perfect score: 1032

Sequence: 1 MKKXGVSIVIGRLNSGDTA.....VAKAVDFIPVESLETTMRSP 197

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-O=/cpn2.1/usptc03pool/US09965594/runat.29082003.151918.28302/app_query.fasta.1.2872
-DB=N_Geneseq 19Jun03 -QEXT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdl
-LIST=45 -LOCAL=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09965594_8CGN_1.1_1412 -runat.29082003.151918.28302 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6
-FCGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq.19Jun03.*

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
25: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB	ID	Description
1	1032	100.0	594	21	AAA73330	Hepatitis C virus
2	1015	98.4	594	21	AAA73331	Hepatitis C virus
3	998	96.7	588	21	AAA73329	Hepatitis C virus
4	993	96.2	594	21	AAA73335	Hepatitis C virus
5	985	95.4	594	21	AAA73332	Hepatitis C virus
6	973	94.3	594	21	AAA73333	Hepatitis C virus
7	963	93.3	594	21	AAA73334	Hepatitis C virus
8	959	92.9	588	21	AAA73328	Hepatitis C virus
9	930.5	90.2	12734	24	ABA95615	Chimeric BVDV/HCV
10	921.5	89.3	612	25	ABX15706	Anti-viral synthet
11	901.5	87.4	5300	10	AA920977	Combined open read
12	901.5	87.4	5360	10	AA920327	Hepatitis C virus
13	901.5	87.4	6905	10	AA921033	Combined open read
14	901.5	87.4	7310	10	AA921066	Combined open read
15	901.5	87.4	7310	10	AA920336	Composite hepatiti
16	901.5	87.4	7310	16	AA098221	Hepatitis C virus
17	901.5	87.4	8136	21	AAA75296	cDNA sequence comp
18	901.5	87.4	9133	20	AA207656	Nucleotide sequenc
19	901.5	87.4	9185	11	AA005956	Sense strand of th
20	901.5	87.4	9185	12	AA010566	Hepatitis C virus
21	901.5	87.4	9185	21	AAA75297	Sense strand of HC
22	901.5	87.4	9400	13	AAQ21144	Compiled HCV cDNA.
23	901.5	87.4	9401	17	AAQ12710	Hepatitis C virus
24	901.5	87.4	9401	18	AA929981	HCV polyprotein co
25	901.5	87.4	9401	19	AAV09989	Hepatitis C virus
26	901.5	87.4	9401	24	AAQ35043	Hepatitis C virus
27	900.5	87.3	9502	15	AAQ74770	Hepatitis C virus
28	899	87.1	549	21	AAA70344	Hepatitis C virus
29	899	87.1	2058	24	AAQ29795	Hepatitis C virus
30	899	87.1	2058	24	ABK15344	HCV-1 NS3/4a mutan
31	899	87.1	2058	25	ABX14410	Hepatitis C virus
32	899	87.1	8145	20	AAQ23259	DNA encoding HCV-1
33	898.5	87.1	9185	20	AAQ26737	Plasmid pET-BS(+)
34	898.5	87.1	9185	20	AAQ00459	Nucleotide sequenc
35	897.5	87.0	8316	11	AAQ05955	Hepatitis C virus
36	897	86.9	1933	20	AAQ23258	Hepatitis C virus
37	896.5	86.9	9646	19	AAV59361	HCV NS3 DNA. Hepa
38	896.5	86.9	9646	24	ABK87285	cDNA encoding hepa
39	896.5	86.9	12980	19	AAV59364	Hepatitis C virus
40	896.5	86.9	12980	24	ABK87286	Hepatitis C virus
41	896.5	86.9	16622	21	AAQ23612	Nucleotide sequenc
42	895	86.7	2061	24	AAQ34500	Hepatitis C virus
43	895	86.7	2061	24	AAQ31767	Hepatitis C virus
44	894.5	86.7	1998	20	AAQ80355	HCV NS4A-NS3 compl
45	892.5	86.5	9365	24	AAQ25518	Hepatitis C virus

ALIGNMENTS

RESULT 1
AAA73330
ID AAA73330 standard; DNA; 594 BP.
XX
XX
AC AAA73330;
XX
XX
DT 19-DEC-2000 (first entry)
DE
DE
XX Hepatitis C virus NS4A-NS3 fusion protease coding sequence #3.
XX
XX Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW liver failure; liver cancer; mutant; muten; ds.
XX
XX Hepatitis C virus.
OS Synthetic.
XX
XX
FH Key Location/Qualifiers

```

FT CDS 1..594
PT /*tag= a
XX /product= "NS4A-NS3 fusion protein #3"
PN WO200040707-A1.
PD 13-JUL-2000.
XX
PF 06-JAN-2000; 2000WO-US00345.
XX
PR 08-JAN-1999; 99US-0115271.
XX
PA (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX WPI; 2000-465976/40.
DR P-PSDB; AAB15221.
XX
XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
PT amino acid, useful for screening inhibitors that may treat hepatitis C
PT
XX
XX Claim 26; Fig 13; 66pp; English.
XX
XX The present sequence is the coding sequence for a mutated version of a
CC fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
CC protease enzymes. These proteins are both essential for the replication
CC of the virus, acting to cleave its replicative proteins from the
CC polypeptide produced from the HCV genome. Inhibitors of the two proteins
CC should be effective as antiviral treatments of HCV infection. This is
CC useful as HCV can lead to chronic liver disease such as cirrhosis, liver
CC failure and liver cancer. The present invention concerns a number of NS3
CC mutants and NS3-NS4A fusion proteins which can be used to identify
CC inhibitors of this type, as well as enabling structural studies of the
CC protease and protease-inhibitor complexes. The protein produced from this
CC sequence contains the alpha-helix0-1 variant.
XX
SQ Sequence 594 BP: 103 A; 186 C; 156 G; 149 T; 0 other:

Alignment Scores:
Pred. No.: 2,42e-85 Length: 594
Score: 1032.00 Matches: 197
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 21 Gaps: 0

US-09-965-594-14 (1-197) x AAA73330 (1-594)
QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
DB 1 ATGAATAAAAGAGATCGGTGTTATCGTCGGCGGTATCAACCTGTCGGTGACCGCT 60
QY 21 TyrAlaGlnGlnThrArgGlyGluGlyCysGlnThrSerGlnThrGlyArgAsp 40
DB 61 TACGCTCAGCAGACTCGAGGTGAGGAGGTGCCAAGAAACCTCCAGACCGGTGCTGAC 120
QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAla 60
DB 121 AAACACAGGTGAAGGTGAAGTTCAGATGTTCCACCGCTGCTCAGACTTCCTGGCT 180
QY 61 ThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
DB 181 ACCGTGATCAGCGGTGTTGCTGACCGCTTACACCGGTGCTGTACCGGTACCATGCT 240
QY 81 SerProLysGlyProValIleGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp 100
DB 241 TCCCGAAGGTCCGGTTATCCAGATGTACACCAACGTTGACAAAGACCTGTTGTTGG 300
QY 101 ProAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerAspLeuTyr 120

```

```

DB 301 CCGGCTCGCAGGGTTCCGTTCCCTGCACCCCGTGCACCTGCGGTTCTCCGACCTGTAC 360
QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
DB 361 CTGGTTACCGCTCAGCTGACGTTATCCCGGTTCCTCGTGTGAGTCCCGTGGTTC 420
QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
DB 421 CTGCTGTCCCGCTCCGATCTCTACCTGAAAGGTTCTCCGGTGGTCCGCTGCTGTC 480
QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys 180
DB 481 CCGGCTGGTCAAGCTGTGGTATCTCCGTGCTGTTGTCACCCGTTGGTGTCTAAA 540
QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
DB 541 GCTGTGTGACTTCATCCCGGTTCGATCCCTGGAAACCAACCATGCTGCTCCCG 591

RESULT 2
AAA73331
ID AAA73331 standard; DNA; 594 BP.
XX
AC AAA73331;
XX
DT 19-DEC-2000 (first entry)
XX
DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #4.
XX
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
XX liver failure; liver cancer; mutant; mutein; ds.
XX
OS Hepatitis C virus.
XX
XX Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1..594 a
FT /tag= a
FT /product= "NS4A-NS3 fusion protein #4"
XX
XX WO200040707-A1.
XX
PN 13-JUL-2000.
PD
XX
PF 06-JAN-2000; 2000WO-US00345.
XX
PR 08-JAN-1999; 99US-0115271.
XX
PA (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX WPI; 2000-465976/40.
DR P-PSDB; AAB15222.
XX
PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
PT amino acid, useful for screening inhibitors that may treat hepatitis C
PT
XX
XX Claim 26; Fig 14; 66pp; English.
XX
XX The present sequence is the coding sequence for a mutated version of a
CC fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
CC protease enzymes. These proteins are both essential for the replication
CC of the virus, acting to cleave its replicative proteins from the
CC polypeptide produced from the HCV genome. Inhibitors of the two proteins
CC should be effective as antiviral treatments of HCV infection. This is
CC useful as HCV can lead to chronic liver disease such as cirrhosis, liver
CC failure and liver cancer. The present invention concerns a number of NS3
CC mutants and NS3-NS4A fusion proteins which can be used to identify
CC inhibitors of this type, as well as enabling structural studies of the
CC protease and protease-inhibitor complexes. The protein produced from this
CC sequence contains the alpha-helix0-1 variant.
XX

```


Alignment Scores:		
Pred. No.:	8.79e-82	594
Score:	993.00	192
Percent Similarity:	97.46%	Conservatives: 0
Best Local Similarity:	97.46%	Mismatches: 5
Query Match:	96.22%	Indels: 0
DB:	21	Gaps: 0
US-09-965-594-14 (1-197)	x	AAA73335 (1-594)

Db 61 TACGCTCAGCAGACTCGAGGTGCTGGTTGCATCATCACCTCCCTGACCGGTGCTGAC 120

QY
41
Db

121

AAAAACGAGTTGAAGCTGAAGTTCAGATCGTTCCACCGCTGCTCAGACCTTCCTGGCT

180

QY 01 nmCys128asgdyvaCys131pimvayrnnisrjafesrjnnmagnnifera 80
|||||
181 ACCTGCATCAACGGTGTTCGTGACGGTGTACACGGTGTGGTACCGTACCATCGCT 240
Db
QY 81 SerProLysGlyproValIleGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr 100

Db 241 TCCCGAAAGGTCGGTTATCCAGATGTACCAACGTTGACAAAGACCTGGTTGGTTGG 300

301	CGGGTCGGCAGGGTTCGGTTCCTGACCCCTGCACCTGCGGGTTCCTCGACCTGAC	360
Db		
Qy	121	LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer

db 361 CTGGTTACCCGTCACGCTGACGTTATCCCGGTTCTCGTCGGTGACTCCCGTGGTTCC 420

141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160

pb 421 CTGCTGTCCCCGGTCCGATCTCCTACCTGAAAGGTTCTCCGGTGGTCCGGCTGCTGTGC 480

161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLeu 180

D_b 481 CCGGCTGGTCA CCGTGTGGTATCTCCGTGGTGTGGTGGTGGTAA 540

181 AlavalAsppheilleprovalGluserLeuGluThrThrMetArgSerPro 197

Db 541 GCTGTTGACTTCATCCCGGTTGAATCCCTGGAACCAACCATGCGTTCCTCCG 591

RESULT 5

ID AAA73332 standard; DNA; 594 BP.

AC AAA73332;

DT 19-DEC-2000 (first entry)

DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #5.
XX
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW liver failure; liver cancer; mutant; mutant; ds.

PN WO200040707-A1.

PD 13-JUL-2000.
 XX
 PF 06-JAN-2000; 2000MO-US00345.
 XX
 PR 08-JAN-1999; 99US-0115271.
 XX (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX
 PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
 XX
 DR WPI: 2000-465976/40.
 DR P-PSDB; AAB15223.
 XX
 PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 PT
 XX
 PS Claim 26; Fig 15; 66pp; English.
 XX
 CC The present sequence is the coding sequence for a mutated version of a
 CC fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
 CC protease enzymes. These proteins are both essential for the replication
 CC of the virus, acting to cleave its replicative proteins from the
 CC polyprotein produced from the HCV genome. Inhibitors of the two proteins
 CC should be effective as antiviral treatments of HCV infection. This is
 CC useful as HCV can lead to chronic liver disease such as cirrhosis, liver
 CC failure and liver cancer. The present invention concerns a number of NS3
 CC mutants and NS3-NS4A fusion proteins which can be used to identify
 CC inhibitors of this type, as well as enabling structural studies of the
 CC protease and protease-inhibitor complexes. The protein produced from this
 CC sequence contains the alpha-helix0-1 variant.
 XX
 SQ Sequence 594 BP; 105 A; 189 C; 153 G; 147 T; 0 other;

Alignment Scores:
 Pred. No.: 4.72e-81 Length: 594
 Score: 985.00 Matches: 191
 Percent Similarity: 96.95% Conservative: 0
 Best Local Similarity: 96.95% Mismatches: 6
 Query Match: 95.45% Indels: 0
 DB: 21 Gaps: 0

US-09-965-594-14 (1-197) x AAA73332 (1-594)

QY 1 MetLysLysGlyGlyValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
 DB 1 ATGAAAAAAGAGATCCGTGTATCTCGCGCGTATCAACCTGTCGGTGACACCGCT 60
 QY 21 TyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlnThrSerGlnThrGlyArgAsp 40
 DB 61 TACGCTCAGCAGACTCGAGGTGAGGAGGTGCGCAAGAAACCTCCACAGCGGTGCTGAC 120
 QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAla 60
 DB 121 AAAAAACAGGTGAAGTGAAGTTTCAGATCGTTTCCACCGCTACCCAGACTTCCTCGCT 180
 QY 61 ThrCysIleAsnGlyValCysTrpThrValTyHisGlyAlaGlyThrArgThrIleAla 80
 DB 181 ACCTCCATCAACGGTGTCTGTGGACGGTTTACACCGTCTGTACCGGTACCATCGCT 240
 QY 81 SerProLysGlyProValIleGlnMetTyThrThrAsnValAspLysAspLeuValGlyTrp 100
 DB 241 TCCCGAAAGTCCGGTTTACCCAGATGTACCAACCGTTGACAAAGACCTGGTTGGTTGG 300
 QY 101 ProIleProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTy 120
 DB 301 CAGGCTCCGACAGGTTCCTCCCTGCTACCCCGTGCACCTGCGGTTCCTCCGACCTGTAC 360
 QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
 DB 361 CTGGTTACCCCGTACCGGTATCCCGGTTCGTCGTGTCGTGTCGTGTCGTGTCGTGTC 420

QY 141 LeuLeuSerProArgProIleSerTyLeuLysGlySerSerGlyGlyProLeuLeuCys 160
 DB 421 CTGCTGTCGCCGCGTCCGATCTCTACCTGAAGGTTCCTCCGGTGGTCCGCTGCTGTC 480
 QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys 180
 DB 481 CCGGCTGGTCAAGCTGTGTGTATCTTCCGTGCTGCTGTTCCACCGGTGCTGCTGCTAAA 540
 QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
 DB 541 GCTGTGACTTCATCCCGGTGGAATCCCTGGAAACCAACCATCGGTTCGCCG 591

RESULT 6

AAA73333
 ID AAA73333 standard; DNA: 594 BP.
 XX
 AC AAA73333;
 XX
 DT 19-DEC-2000 (first entry)
 XX
 DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #6.
 XX
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
 KW liver failure; liver cancer; mutant; muten; ds.
 XX
 OS Hepatitis C virus.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..594 /tag= a
 FT /product= "NS4A-NS3 fusion protein #6"
 FT
 XX WO200040707-A1.
 PN
 XX 13-JUL-2000.
 PD
 XX 06-JAN-2000; 2000MO-US00345.
 PF
 XX 08-JAN-1999; 99US-0115271.
 PR
 XX (BRIM) BRISTOL-MYERS SQUIBB CO.
 PA
 XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
 PI
 XX WPI: 2000-465976/40.
 DR P-PSDB; AAB15224.
 DR
 XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 PT
 XX
 PS Claim 26; Fig 16; 66pp; English.
 XX
 CC The present sequence is the coding sequence for a mutated version of a
 CC fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
 CC protease enzymes. These proteins are both essential for the replication
 CC of the virus, acting to cleave its replicative proteins from the
 CC polyprotein produced from the HCV genome. Inhibitors of the two proteins
 CC should be effective as antiviral treatments of HCV infection. This is
 CC useful as HCV can lead to chronic liver disease such as cirrhosis, liver
 CC failure and liver cancer. The present invention concerns a number of NS3
 CC mutants and NS3-NS4A fusion proteins which can be used to identify
 CC inhibitors of this type, as well as enabling structural studies of the
 CC protease and protease-inhibitor complexes. The protein produced from this
 CC sequence contains the alpha-helix0-7 variant.
 XX
 SQ Sequence 594 BP; 104 A; 191 C; 152 G; 147 T; 0 other;

Alignment Scores:
 Pred. No.: 5.88e-80 Length: 594
 Score: 973.00 Matches: 188

```

Percent Similarity: 96.45% Conservative: 2
Best Local Similarity: 95.43% Mismatches: 7
Query Match: 94.28% Indels: 0
DB: 21 Gaps: 0

US-09-965-594-14 (1-197) x AAA73333 (1-594)

QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
DB 1 ATGAAATAAAGGATCCGTTGTTATCGCGCCGTATCAACCTGTCGGTGACACCGCT 60

QY 21 TyrAlaGlnGlnThrArgGlyGluGlyCysGlnGluThrSerGlnThrGlyArgAsp 40
DB 61 TAGCGTCAGCAGACGTCGAGGTGAGGAGGTGCCAGAAAGACCTCCACACCGGTGCTGAC 120

QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAla 60
DB 121 AAAAACCCAGGTTGAAGGTGAAGTTTCCAGATCGTTCCACCGCTACCCAGACCTTCCTGGCT 180

QY 61 ThrCysIleAsnGlyValCysTyrThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
DB 181 ACCTCCATCAACGGTCTCTGTGGACCGTTTACCACGGTGTGTTACCGTACCATCGCT 240

QY 81 SerProLysGlyProValIleGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr 100
DB 241 TCCCGGAAAGGTCCGGTTACCCAGATGTACACCAACGTTGACAAAGACCTGGTGGTTGG 300

QY 101 ProAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
DB 301 CAGGCTCCGACGGTCCGTTCCCTGACCCCGTGCACCTGCGGTTCCTCCGACCTGTAC 360

QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
DB 361 CTGGTGTACCGCTACCGTGACGTTATCCCGTTCTGCTGCTGCTGCTGCTGCTGCTGCT 420

QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
DB 421 CTGCTGTCCCGGTCGGATCTCCACCTGAAAGGTTCCTCCGGTGGTCCGCTGCTGTC 480

QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys 180
DB 481 CCGGCTGGTCAGCTGTTGTTATCTCCGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540

QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
DB 541 GCTGTGTACTTCATCCCGTTGAATCCCTCGAAACACCATCGTTCCTCCCG 591

RESULT 7
AAA73334
ID AAA73334 standard; DNR; 594 BP.
XX
AC AAA73334;
XX
XX 19-DEC-2000 (first entry)
XX
DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #7.
XX
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW liver failure; liver cancer; mutant; mutin; ds.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
Key Location/Qualifiers
FT CDS 1..594
FT /tag= a
FT /product= "NS4A-NS3 fusion protein #7"
XX
XX WO200040707-A1.
XX
XX 13-JUL-2000.
XX
XX 06-JAN-2000; 2000WO-0500345.

```

```

XX
PR 08-JAN-1999; 99US-0115271.
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX
XX WPI: 2000-465976/40.
XX P-PSDB; AAB15225.
XX
PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
PT amino acid, useful for screening inhibitors that may treat hepatitis C
XX
XX Claim 26; Fig 17; 66pp; English.
XX
CC The present sequence is the coding sequence for a mutated version of a
CC fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
CC protease enzymes. These proteins are both essential for the replication
CC of the virus, acting to cleave its replicative proteins from the
CC polypeptide produced from the HCV genome. Inhibitors of the two proteins
CC should be effective as antiviral treatments of HCV infection. This is
CC useful as HCV can lead to chronic liver disease such as cirrhosis, liver
CC failure and liver cancer. The present invention concerns a number of NS3
CC mutants and NS3-NS4A fusion proteins which can be used to identify
CC inhibitors of this type, as well as enabling structural studies of the
CC protease and protease-inhibitor complexes. The protein produced from this
CC sequence contains the alpha-helix0-7 variant.
XX
SQ Sequence 594 BP; 105 A; 192 C; 151 G; 146 T; 0 other;

Alignment Scores:
Pred. No.: 4.82e-79 Length: 594
Score: 963.00 Matches: 187
Percent Similarity: 95.94% Conservative: 2
Best Local Similarity: 94.92% Mismatches: 8
Query Match: 93.31% Indels: 0
DB: 21 Gaps: 0

US-09-965-594-14 (1-197) x AAA73334 (1-594)

QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
DB 1 ATGAAATAAAGGATCCGTTGTTATCGCGCCGTATCAACCTGTCGGTGACACCGCT 60

QY 21 TyrAlaGlnGlnThrArgGlyGluGlyCysGlnGluThrSerGlnThrGlyArgAsp 40
DB 61 TAGCGTCAGCAGACGTCGAGGTGAGGAGGTGCCAGAAAGACCTCCACACCGGTGCTGAC 120

QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAla 60
DB 121 AAAAACCCAGGTTGAAGGTGAAGTTTCCAGATCGTTTCCACCGCTACCCAGACCTTCCTGGCT 180

QY 61 ThrCysIleAsnGlyValCysTyrThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
DB 181 ACCTCCATCAACGGTGTCTGTGGACCGTTTACCACGGTGTGTTACCGTACCATCGCT 240

QY 81 SerProLysGlyProValIleGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr 100
DB 241 TCCCGGAAAGGTCCGGTTACCCAGATGTACACCAACGTTGACAAAGACCTGGTGGTTGG 300

QY 101 ProAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
DB 301 CAGGCTCCGACGGTCCGTTCCCTGACCCCGTGCACCTGCGGTTCCTCCGACCTGTAC 360

QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
DB 361 CTGGTGTACCGCTACCGTGACGTTATCCCGTTCTGCTGCTGCTGCTGCTGCTGCTGCT 420

QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
DB 421 CTGCTGTCCCGGTCGGATCTCCACCTGAAAGGTTCCTCCGGTGGTCCGCTGCTGTC 480

```

QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys 180
 Db 481 CCGCGTGGTCACCGTGTGTGATCTCCGTCGCTGCTGTTCCACCCGCGTGTGTGCTAAA 540
 QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
 Db 541 GCTGTTGACTTCATCCCGGTTGATCCTGGAACCCATGCGTTCCTCCCG 591

RESULT 8
 AAA73328
 ID AAA73328 standard; DNA; 588 BP.
 AC AAA73328;
 XX
 XX
 DT 19-DEC-2000 (first entry)
 DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #1.
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
 KW liver failure; liver cancer; ds.
 XX Hepatitis C virus.
 OS Synthetic.
 XX
 XX Key Location/Qualifiers
 FT CDS 1..588
 FT /*tag= a
 FT /product= "NS3-NS4A fusion protein"
 XX
 PN W0200040707-A1.
 XX
 XX 13-JUL-2000.
 XX
 XX 06-JAN-2000; 2000WO-US00345.
 XX
 XX 08-JAN-1999; 99US-0115271.
 XX
 XX (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX
 XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
 XX WPI; 2000-465976/40.
 DR P-PSDB; AAB15212.
 XX
 XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 PT
 XX
 XX Disclosure; Fig 10; 66pp; English.
 XX
 CC The present sequence is the coding sequence for a fusion protein created
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
 CC proteins are both essential for the replication of the virus, acting to
 CC cleave its replicative proteins from the polyprotein produced from the
 CC HCV genome. Inhibitors of the two proteins should be effective as
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A
 CC fusion proteins which can be used to identify inhibitors of this type, as
 CC well as enabling structural studies of the protease and
 CC protease:inhibitor complexes.
 XX
 SQ Sequence 588 BP; 97 A; 183 C; 153 G; 155 T; 0 other;

Alignment Scores:

Pred. No.: 1,1e-78 Length: 588
 Score: 959.00 Matches: 188
 Percent Similarity: 95.94% Conservative: 1
 Best Local Similarity: 95.43% Mismatches: 6
 Query Match: 92.93% Indels: 2
 DB: 21 Gaps: 1

US-09-965-594-14 (1-197) x AAA73328 (1-588)
 QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
 Db 1 ATGAAAAAAGAGTTCCGTTGTTATCTGCGCCGTATAGTACTAGACGGT-----GCT 54
 QY 21 TyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyAlaGAsp 40
 Db 55 TAGCTCAGCAGACTCGAGGTCTGCTGGGTGTCATCATCCTCCCTGACCGGTCTGTGAC 114
 QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAla 60
 Db 115 AAAAACCCAGGTTGAAGGTGAAGTTCAGATGCTTCCACCGCTGCTCAGACCTTCCTGGCT 174
 QY 61 ThrCysIleAsnGlyValCysTyrThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
 Db 175 ACCTGCATCAACGGTGTGTTGCTGGACCGTTTACCACGGTCTGTTACCGGTACCATCGCT 234
 QY 81 SerProLysGlyProValIleGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr 100
 Db 235 TCCCCGNAAGTCCGGTTATCCAGATGTACACCAACGTTGACAAAGACCTGTTGGTTGG 294
 QY 101 ProAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
 Db 295 CCGGCTCCGCAGGGTTCCTGCTGACCCCGTGCACCTGCGGTTCCTCCGACCTGTAC 354
 QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
 Db 355 CTGGTTACCCGTCACGTCAGCTTATCCCGGTTTCGTCGTCGTGACTCCCGTGGTTCC 414
 QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
 Db 415 CTGCTGTCCCGCGTCCGATCTCTACCTGAAGGTTCTCCCGTGGTCCGCTGCTGTC 474
 QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys 180
 Db 475 CCGGCTGGTCACGCTGTTGTTATCTTCCGTCGCTGTTTGCACCCGCTGTTGCTAAA 534
 QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
 Db 535 GCTGTTGACTTCATCCCGGTTGAATCCCTGGAACCAACCATGCGTTCCTCCCG 585

RESULT 9
 ABA95615
 ID ABA95615 standard; DNA; 12734 BP.
 XX
 AC ABA95615;
 XX
 DT 21-MAR-2002 (first entry)
 XX
 DE Chimeric BVDV/HCV NS3-wt sequence.
 XX
 KW Pestivirus; Npro; protease; NS3; screening; ds.
 OS Chimeric - Bovine viral diarrhea virus.
 OS Chimeric - Hepatitis C virus.
 XX
 PN US6326137-B1.
 XX
 PD 04-DEC-2001.
 XX
 PF 25-JUN-1999; 99US-0344456.
 XX
 PR 25-JUN-1999; 99US-0344456.
 XX
 PA (SCHE) SCHERING CORP.
 XX
 PI Hong Z, Lai VCH, Lau JYN;
 XX
 DR WPI; 2002-121103/16.
 XX
 PT Nucleic acid construct encoding chimeric Hepatitis C Virus (HCV)

PT pestivirus genome where the Npro protease gene is replaced with NS3
PT protease gene, useful for in vivo screening of compounds which inhibit
PT HCV infection -
XX
XX Example 2; Columns 17-28; 20pp; English.
XX
CC The present invention relates to a nucleic acid construct encoding a
CC chimeric Hepatitis C virus (HCV)-pestivirus genome. The construct
CC comprises a pestivirus genome where a Npro pestivirus protease gene is
CC replaced with a gene encoding a functional HCV NS3 protease. Furthermore,
CC each junction site recognised by the Npro protease is replaced with a
CC junction site recognised by the HCV NS3 protease. The construct is useful
CC for screening compounds that inhibit HCV in vivo by inhibiting HCV
CC protease, where screening may be in cell culture or in an animal model.
CC The present sequence is a chimeric clone of BVDV (bovine viral diarrhoea
CC virus)/HCV NS3-wt, which was used to illustrate the present invention.
XX
SQ Sequence 12734 BP; 4032 A; 2604 C; 3295 G; 2803 T; 0 other;

Alignment Scores:
Pred. No.: 1.94e-74 Length: 12734
Score: 930.50 Matches: 181
Percent Similarity: 94.87% Conservative: 4
Best Local Similarity: 92.82% Mismatches: 7
Query Match: 90.16% Indels: 3
DB: 24 Gaps: 1

US-09-965-594-14 (1-197) x ABA95615 (1-12734)
QY 5 GlySerValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
Db 413 GSTAGTGTGTTTATGTTGTTAGTAATGTTTATCTGTGTAGTATCATCGGGGTAC 472
QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGlnThrSerGlnThrGlyArgAspLys 41
Db 473 GCCACGACGAGAGAGCCCTCTAGGGTGAAGATCACCAGTCTGACTGGCGGGACAAA 532
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAlaThr 61
Db 533 AACCAAGTGGAGGTGAGTCCAGATCGTGTCAACTGCTACCCAAACCTTCCTGGCAACG 592
QY 62 CysIleAsnGlyValCysTrpThrValTyrHisGlyValGlyThrArgThrIleAlaSer 81
Db 593 TGCATCAATGGGTATGCTGGACTGTCTACACGGGGCCGGAACGAGGACCATCGCATCA 652
QY 82 ProGlyGlyProValIleGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpPro 101
Db 653 CCCAAGGGTCTCTCATCCAGATGTATACCAATGTGGACCAAGACCTGTGGGCTGGCCC 712
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
Db 713 GCTCCTCAAGGTTCGGCTCATTCACACCTCGACCTCGCGGCTCCCTCGGACCTTTACCTG 772
QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
Db 773 GTACGAGGACGCGGACGTCATTCCTCGCGCGGAGGTGATACGAGGGTAGCCTG 832
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
Db 833 CTTTGGCCCGGCCCCATTCTCTACCTAAAGGCTCTCTCGGGGGTCCGCGTGTGTGCCCC 892
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181
Db 893 GCGGAGACCGCTGGGCGCTATTTCAGGGCCGCGGTGTGCACCGGTGGAGTGGCAAGGCG 952
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196
Db 953 GTGGACTTTATCCCTGTGGAGAACCTAGACACCAACCATGATGCC 997

RESULT 10
ID ABX15706
XX US-09-965-594-14 (1-197) x ABX15706 (1-612)

AC ABX15706;
XX 28-MAR-2003 (first entry)
XX
XX Anti-viral synthetic prototoxophore associated DNA sequence.
DE
XX Hepatitis C; ds; viral prototoxophore; anti-viral; tumour;
KW virus; infection; antitumour; toxophore; human immunodeficiency virus;
KW HIV infection; herpes simplex virus; HSV; rhinovirus; NS3 protease.
XX
OS Unidentified.
XX WO200287500-A2.
XX 07-NOV-2002.
XX 26-APR-2002; 2002WO-US13223.
XX 27-APR-2001; 2001US-286893P.
XX (NEWB-) NEWBIOTICS INC.
XX Cathers BE, Neuteboom STC, Shepard HM;
XX WPI: 2003-167102/16.
XX
PT Novel synthetic viral prototoxophore for treating viral infections, has
PT toxin moiety incorporated into substrate domain specific for viral
PT enzyme, bound and modified by viral enzyme to get converted into
PT toxophore -
XX
XX Example 1; Page 62; 66pp; English.
XX This invention relates to a novel synthetic viral prototoxophore
XX comprising a toxin moiety operatively incorporated into a substrate
XX domain specific for a viral enzyme. This prototoxophore may be bound
XX and modified by the viral enzyme thus converting it to a toxophore.
XX Also disclosed in the invention is a method for enhancing the anti-viral
XX effect of an antiviral agent, this method comprises contacting a cell,
XX infected with a virus or is susceptible to infection, with a
XX prototoxophore. The invention further comprises an assay to identify
XX anti-viral agents, comprising contacting an infected cell with a
XX candidate agent and comparing the ability of the agent to inhibit the
XX growth or infectivity of the virus in the cell. The prototoxophores
XX of the invention may have virucide or antitumour activity. The
XX prototoxophores of the invention may be useful for reducing or
XX inhibiting viral infectivity, by contacting a cell (e.g. lymphocyte,
XX nerve cell, connective tissue cell, muscle cell or hepatocyte) which
XX is infected with a virus or is susceptible to infection with a virus, with
XX an effective amount of the prototoxophore. The cells are cell lines
XX adapted to long term continuous culture or isolated from a subject.
XX The prototoxophore is also useful for ameliorating the severity of a
XX viral infection in a subject, where the virus is selected from human
XX immunodeficiency virus (HIV), herpes simplex virus (HSV), rhinovirus and
XX hepatitis virus, by administering an effective amount of the
XX prototoxophore to the subject. The prototoxophores of the invention are
XX also useful for treating tumours. The present sequence represents an
XX antiviral prototoxophore associated DNA sequence, this sequence is
XX described as a recombinant NS3/NS4 fusion protein in example 1 of
XX the invention although it is clearly not a protein sequence.
SQ Sequence 612 BP; 120 A; 171 C; 191 G; 130 T; 0 other;

Alignment Scores:
Pred. No.: 3.07e-75 Length: 612
Score: 921.50 Matches: 181
Percent Similarity: 94.36% Conservative: 3
Best Local Similarity: 92.82% Mismatches: 8
Query Match: 89.29% Indels: 3
DB: 25 Gaps: 1

US-09-965-594-14 (1-197) x ABX15706 (1-612)

Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
 Db 19 GGTAGTGGTCATTGGGTAGGATCATTTTGTCCGGTAGTGTAGTATACGGCGTAC 78
 Qy 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41
 Db 79 GCCCAGCAGACAAGGGCCCTCTAGGTGCTATATCAATCAACAGCCCTAACTGGCCGGGACAAA 138
 Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAlaThr 61
 Db 139 AACCAAGTGGAGGTGAGGTCCAGATTGTGTCACTGCTGCCCAACCTTCTCGGCACGC 198
 Qy 62 CysIleAsnGlyValCysTrrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
 Db 199 TGCATCAATGGGTGTGTGACTGTCTACACGGGGCCGGAACGAGGACCATCGCGTCA 258
 Qy 82 ProlGlyProValIleGlnMetTyrThrAsnValAspLysAspLeuValGlyTrrpPro 101
 Db 259 CCCAAGGGTCTGTCTCCAGATGTATACCAATGTAGACCAAGACCTTGTGGGCTGGCCC 318
 Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
 Db 319 GCTTCGCAAGGTACCGCTCATTCACACCTGCACTTGGCGCTCTCGGACCTTTACTGT 378
 Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
 Db 379 GTCACGAGGCACGCGATGTCTATCCGCTGCGCGCGGGGTGTATAGCAGGGGCGCCTG 438
 Qy 142 LeuSerProArgProIleSerTyrLeuIleGlySerSerGlyGlyProLeuLeuCysPro 161
 Db 439 CTGTGCGCCCGCCCATTTCTTCTTAAAGGCTCTCGGGGGGTCTCGCTGTGTGGCCCC 498
 Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181
 Db 499 GCGGGCAGCGCGTGGCATATTTAGGCGCGCGTGTGCACCGCTGGAGTGGCTAAGCG 558
 Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196
 Db 559 GTGACTTTATCTCTGTGGAGAACCTAGAGACAACCATGAGGTCC 603
 RESULT 11
 AAN92097
 ID AAN92097 standard; DNA; 5300 BP.
 XX
 AC AAN92097;
 DT 25-MAR-2003 (updated)
 DT 02-MAR-1990 (first entry)
 XX
 DE Combined open reading frames of the hepatitis C virus (HCV) cDNA in
 DE clones 141 11b, 7f, 7e, 8h, 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c, 14c,
 DE 8f, 33f, 33g and 39c.
 XX
 KW Hepatitis C virus; HCV; non-A, non-B hepatitis; NANBH.
 XX
 OS Hepatitis C virus.
 XX
 FH key Location/Qualifiers
 FT CDS 3..5300
 FT /*tag= a
 XX
 PN EP318216-A.
 XX
 PD 31-MAY-1989.
 XX
 PF 18-NOV-1988; 88EP-0310922.
 XX
 PR 18-NOV-1987; 87US-0122714.
 PR 30-DEC-1987; 87US-0139886.
 PR 26-FEB-1988; 88US-0161072.
 PR 06-MAY-1988; 88US-0191263.
 PR 26-OCT-1988; 88US-0263584.
 PR 14-NOV-1988; 88US-0271450.

XX
 PA
 XX (CHIR) CHIRON CORP.
 PI Houghton M, Choo QL, Kuo G;
 XX
 DR WP1; 1989-159274/22.
 DR P-PSDB; AAP92041.
 XX
 PT Purified hepatitis C virus
 PT - and associated nucleic acids and polypeptide(s)
 XX
 PS Claim 3; Figure 26-1, 26-2, 26-3, 26-4, 26-5, 26-6; 139pp; English.
 XX
 CC It is a double-stranded nucleotide sequence of the open reading frame
 CC (ORF) (tag a) extending through clones 141, 11b, 7f, 7e, 8h 33c, 40b, C
 CC 37b, 35, 36, 81, 32, 33b, 25c, 14c, 8f, 33f, 33g and 39c of hepatitis C
 CC virus (HCV) cDNA. In creating the composite sequence the following
 CC heterogeneities were considered. Clone 33c contains a sequence
 CC of 800 base pairs which overlaps the cDNAs in clones 40b and 37c. In
 CC clone 33c, as well as in 5 other overlapping clones, nucleotide #789 is
 CC a G. However, in clone 37b the corresponding nucleotide is an A. This
 CC heterogeneity may have important ramifications for protein folding.
 CC Nucleotide #2 in clone 8h is a T which may represent a cloning artifact
 CC because the corresponding residue in clone 7e and in 3 other overlapping
 CC clones is an A. Therefore the residue in this position is designated as
 CC an A. The 3'-terminal nucleotide in clone 8f is represented as a T
 CC than a G because the corresponding residue in clone 33f and in 2 other
 CC overlapping clones is a T. The 3' terminal sequence of clone 33f is
 CC represented as ATTC, as is found in the corresponding sequence in clone
 CC 33g and in 2 other overlapping clones, rather than as TTGC, as is found
 CC in clone 33f. Residue #4 in clone 33g is designated an A rather than a T
 CC because the corresponding residue in clone 33f and 2 other overlapping
 CC clones is an A. The 3'-terminus of clone 141 is depicted as TA rather
 CC than AA because the corresponding dinucleotide in clone 11b and 3 other
 CC clones is TA. Potential cloning artifacts have been omitted and instead
 CC the corresponding sequences in non-5'-terminal regions of multiple
 CC overlapping clones are shown. AAN92097 could be used as a source of
 CC oligomeric DNA hybridisation probes to detect the presence of HCV
 CC nucleic acids in samples. The polypeptide(s) it encodes could be used as
 CC immuno- assay reagents and vaccines and to generate antibodies useful in
 CC diagnosis and passive immunotherapy for HCV infection/non-A, non-B
 CC hepatitis.
 CC
 CC (Updated on 25-MAR-2003 to correct PR field.)
 CC (Updated on 25-MAR-2003 to correct PI field.)
 XX
 SO Sequence 5300 BP; 1047 A; 1606 C; 1515 G; 1130 T; 2 other;
 Alignment Scores:
 Pred. No.: 2.93e-72 Length: 5300
 Score: 901.50 Matches: 175
 Percent Similarity: 90.20% Conservative: 9
 Best Local Similarity: 85.78% Mismatches: 11
 Query Match: 87.35% Indels: 9
 DB: 10 Gaps: 1
 US-09-965-594-14 (1-197) x AAN92097 (1-5300)
 Qy 3 LysLysGlySerValValIleValGlyArgIleAsn----- 14
 Db 867 CGCAGGGCGCGGAGATGATGCTCGCGCGCGCGGCGGATGCTCCAGGGGGTGAGG 926
 Qy 15 ---LeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGluCysGlnGlu 33
 Db 927 TTGCTGGCGCCCATCATCGCGGTACGCCAGCAGACAAAGGGGCTCTCTAGGTGCATAATC 986
 Qy 34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 53
 Db 987 ACCAGCCTAACTGGCGGGGACAAAACCAAGTGGAGGTGAGGTCCAGATTGTGCAACT 1046
 Qy 54 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrrpThrValTyrHisGly 73
 Db 1047 GCTGCCCAACACCTTCTCTGGCAACGTCATCAATGGGGTGTGTGGACTGTCTACCAAGG 1106

QY 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 93
 DB 1107 GCGGAACAGGACCATCGTCACCAAGGCTCCTGTTCATCCAGATGTATACCAATGTA 1166
 QY 94 AspLysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThr 113
 DB 1167 GACCAAGACCTTGTGGCTGGCCCGCTCCGCAAGGTAGCCGCTCATTTGACACCCCTGCCT 1226
 QY 114 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 133
 DB 1227 TCGGGCTCTCGGACCTTACCTGTGTACAGGACGACGCGATGTATTCCTCGCGCGG 1286
 QY 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 153
 DB 1287 CGGGGTGATAGCAGGGCAGCCTGTGTGCGCCCGGCCCATTTCTACTTGAAGGCTCC 1346
 QY 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaVal 173
 DB 1347 TCGGGGGGTCCGCTGTGTGCCCCCGGGGACGCGGTGGGCATATTTAGGCGCCGCGTG 1406
 QY 174 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 193
 DB 1407 TGCACCCGTGGAGTGGCTAAGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACACC 1466
 QY 194 MetArgSerPro 197
 DB 1467 ATGAGTCCCCG 1478

RESULT 12
 AAN90327
 ID AAN90327 standard; cDNA; 5360 BP.
 AC AAN90327;
 XX
 DT 25-MAR-2003 (updated)
 DT 11-NOV-1989 (first entry)
 XX
 DE Hepatitis C virus composite probe.
 XX
 KW Hepatitis C virus; composite cDNA; probe; vaccine.
 XX
 OS Pan troglodytes.
 XX
 FH Key Location/Qualifiers
 FT CDS 3..5360
 FT /*tag- a
 XX
 PN GB2212511-A.
 XX
 PD 26-JUL-1989.
 XX
 PF 18-NOV-1988; 88GB-0027024.
 XX
 PR 18-NOV-1987; 87US-0122714.
 PR 30-DEC-1987; 87US-0139486.
 PR 26-FEB-1988; 88US-0161072.
 PR 26-OCT-1988; 88US-0263584.
 XX
 PA (CHIR) CHIRON CORPORATION.
 XX
 PI Houghton M, Choo QL, Kuo G;
 XX
 DR WPI; 1989-215054/30.
 XX
 PT Hepatitis C virus gene - used for prodn. of polynucleotide probes,
 PT polypeptide(s) and antibodies for diagnosis, prevention and treatment
 PT of infection.
 PS Disclosure; Fig. 26; 174pp; English.
 XX
 CC The sequence shows the composite cDNA sequence derived from the aligned
 CC hepatitis C virus (HCV) cDNA's in clones 14i, 11b, 7f, 7e, 8h, 33c, 40b,
 CC 37b, 35, 36, 8i, 32, 33b, 25c, 14c, 8f, 33f, 33g and 39c. The cDNA

CC encodes antigens which react with antibodies in patients with non-A
 CC non-B hepatitis (NANBH). The cDNA can be used to design probes, or to
 CC synthesise polypeptides, which are used to diagnose HCV-induced NANBH,
 CC to raise antibodies for immunoassay or treatment, or to produce
 CC vaccines. See also AAP90158, AAN90303-26, and AAN90328-36.
 CC (Updated on 25-MAR-2003 to correct PR field.)
 XX
 SQ Sequence 5360 BP; 1060 A; 1622 C; 1532 G; 1145 T; 1 other;
 Alignment Scores:
 Pred. No.: 2.97e-72 Length: 5360
 Score: 901.50 Matches: 175
 Percent Similarity: 90.20% Conservative: 9
 Best Local Similarity: 85.78% Mismatches: 11
 Query Match: 87.35% Indels: 9
 DB: 10 Gaps: 1

US-09-965-594-14 (1-197) x AAN90327 (1-5360)

QY 3 LysLysGlySerValValIleValGlyArgIleAsn----- 14
 DB 867 CGCAGGGCCGGAGATACTCTCGGGCCAGCCGATGGATGGTCTCCAAGGGGTGGAGG 926
 QY 15 ---LeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGlyCysGlnGlu 33
 DB 927 TTGCTGGCGCCCATCACGGCTAGCCCGCAGCAGACAAGGGGCCCTCCTAGGGTGCATAATC 986
 QY 34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 53
 DB 987 ACCAGCTTAAGTGGCGGGACAAAACCAAGTGGAGGTGAGGTCCAGATTGTCTCAACT 1046
 QY 54 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysIleThrValThrValHisGly 73
 DB 1047 GCTGCCAAACCTTCTCTGGCAACGTGCATCAATGGGGTGTCTGGACTGTCTACACGGG 1106
 QY 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 93
 DB 1107 GCGGAACGAGGACCATCGCTACCCCAAGGTCCTGTCTCCAGATGTATACCAATGTA 1166
 QY 94 AspLysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThr 113
 DB 1167 GACCAAGACCTTGTGGCTGGCCCGCTCCGCAAGGTAGCCGCTCATTCACACCCCTGCCT 1226
 QY 114 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 133
 DB 1227 TCGGGCTCTCGGACCTTACCTGTGTACAGGACGACGCGCATGTCTCCGCTGCGCGG 1286
 QY 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 153
 DB 1287 CGGGGTGATAGCAGGGGACGCTGTGTGCCCCCGGCCCATTTCTACTTGAAGGCTCC 1346
 QY 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaVal 173
 DB 1347 TCGGGGGGTCCGCTGTGTGCCCCCGGGGACGCGGTGGGCATATTTAGGCGCCGCGTG 1406
 QY 174 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 193
 DB 1407 TGCACCCGTGGAGTGGCTAAGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACACC 1466
 QY 194 MetArgSerPro 197
 DB 1467 ATGAGTCCCCG 1478

RESULT 13
 AAN92103
 ID AAN92103 standard; DNA; 6905 BP.
 XX
 AC AAN92103;
 XX
 DT 25-MAR-2003 (updated)
 DT 02-MAR-1990 (first entry)
 XX
 DE Combined open reading frames of the hepatitis C virus (HCV) cDNAs from

```

DE clones 12f through 15e.
XX Hepatitis C virus; HCV; non-A, non-B hepatitis; NANBH.
KW Hepatitis C virus.
OS
XX
XX Key Location/Qualifiers
FT CDS 3..6905
FT /*tag= a
XX
XX EP318216-A.
XX
XX 31-MAY-1989.
XX
XX 18-NOV-1988; 88EP-0310922.
XX
XX 18-NOV-1987; 87US-0122714.
XX 30-DEC-1987; 87US-0139886.
XX 26-FEB-1988; 88US-0161072.
XX 06-MAY-1988; 88US-0191263.
XX 26-OCT-1988; 88US-0263584.
XX 14-NOV-1988; 88US-0271450.
XX
XX (CHIR ) CHIRON CORP.
XX
XX Houghton M, Choo QL, Kuo G;
XX WPI: 1989-159274/22.
XX P-PSDB; AAP92047.
XX
XX Purified hepatitis C virus
PT - and associated nucleic acids and polypeptide(s)
XX
XX Claim 3; Figure 32-1 - 32-7; 139pp; English.
XX
XX It is a double-stranded nucleotide sequence of the open reading frame
CC (ORF) (tag a) extending through clones 12f to 15e of hepatitis C virus
CC (HCV) cDNA. It can be used to make oligomeric DNA hybridisation probes to
CC detect the presence of HCV nucleic acids in samples. The polypeptide(s)
CC it encodes could be used as immunoassay reagents and vaccines and to
CC generate antibodies useful in diagnosis and passive immunotherapy for
CC HCV infection/non-A, non-B hepatitis.
CC
CC (updated on 25-MAR-2003 to correct PR field.)
CC (updated on 25-MAR-2003 to correct PI field.)
XX
XX Sequence 6905 BP; 1421 A; 2082 C; 1946 G; 1456 T; 0 other;
SQ
XX
XX Alignment Scores:
XX Pred. No.: 4.05e-72 Length: 6905
XX Score: 901.50 Matches: 175
XX Percent Similarity: 90.20% Conservative: 9
XX Best Local Similarity: 85.78% Mismatches: 11
XX Query Match: 87.35% Indels: 9
XX DB: 10 Gaps: 1
XX
XX US-09-965-594-14 (1-197) x AAN92103 (1-6905)
XX
XX 3 LysLysGlySerValValIleValGlyArgIleAsn----- 14
XX
XX 1140 CGCAGGGGGGGGAGATCTGCTGGGCGCAGCGATGGTCTCCAAAGGGGTGGAGG 1199
XX
XX 15 ---LeuSerGlyAspThrAlaThrAlaGlnGlnThrArgGlyGluGlyCysGlnGlu 33
XX
XX 1200 TTGCTGGGCCCATCAGCGGTACGCCAGCAGCAGCAAGGGGCTCTAGGGTGCAATC 1259
XX
XX 34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyValGlnIleValSerThr 53
XX
XX 1260 ACCAGCTACTTGGCGGGGCAAAACCAAGTGGAGGTGAGGTCCAGATTGTGTCAACT 1319
XX
XX 54 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysThrPheValThrHisGly 73
XX
XX 1320 GCTGCCCAACCTTCTCTGGCAAGTGCATCAATGGGGTGTGCTGGACTGTCTACCACGG 1379
XX

```

```

QY 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 93
DB 1380 GCCGGAACGAGGACCATCGCTCACCACCAAGGTCTCTGTCCATCCAGATGTATACCAATGTA 1439
QY 94 AspLysAspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThr 113
DB 1440 GACCAAGACCTTGTGGGCTGGCCGCTCCGCAAGGTAGCGCTCATTTGACACCTGCACT 1499
QY 114 CysGlySerSerAspLeuThrValThrArgHisAlaAspValIleProValArgArg 133
DB 1500 TCGGCTCTCTGGACCTTTACTGTGTACGAGGACGCCGATGATTCCTCTTCAAGGCTCC 1559
QY 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 153
DB 1560 CGGGGTGATAGCAGGGGCGAGCTGTGTGCGCCCGGCCCATTTCTCTACTTCAAGGCTCC 1619
QY 154 SerGlyGlyProLeuLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 173
DB 1620 TCGGGGGGTCTGCTGTGTGCGCCGCGGGGACGCCGTGGGCATATTTAGGGCGCGGTG 1679
QY 174 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 193
DB 1680 TGCACCCGTGGAGTGGCTAAGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACC 1739
QY 194 MetArgSerPro 197
DB 1740 ATGAGTCCCGC 1751
RESULT 14
AAN92106
ID AAN92106 standard; DNA; 7310 BP.
XX
XX AAN92106;
XX
XX 25-MAR-2003 (updated)
XX 02-MAR-1990 (first entry)
XX
XX Combined open reading frames of the hepatitis C virus (HCV) cDNAs from
DE clones K9-1 through 15e.
XX
XX Hepatitis C virus; HCV; non-A, non-B hepatitis; NANBH.
XX
XX Hepatitis C virus.
XX Key Location/Qualifiers
XX CDS 3..7310
XX /*tag= a
XX
XX EP318216-A.
XX
XX 31-MAY-1989.
XX
XX 18-NOV-1988; 88EP-0310922.
XX
XX 18-NOV-1987; 87US-0122714.
XX 30-DEC-1987; 87US-0139886.
XX 26-FEB-1988; 88US-0161072.
XX 06-MAY-1988; 88US-0191263.
XX 26-OCT-1988; 88US-0263584.
XX 14-NOV-1988; 88US-0271450.
XX
XX (CHIR ) CHIRON CORP.
XX
XX Houghton M, Choo QL, Kuo G;
XX
XX WPI: 1989-159274/22.
XX P-PSDB; AAP92050.
XX
XX Purified hepatitis C virus
XX - and associated nucleic acids and polypeptide(s)
XX
XX Claim 3; Figure 47-1 - 47-8; 139pp; English.
XX

```

CC It is a double-stranded nucleotide sequence of the open reading frame
 CC (ORF) (tag a) extending through clones K9-1 to 15c of hepatitis C virus
 CC (HCV) cDNA. It can be used to make oligomeric DNA hybridisation probes to
 CC detect the presence of HCV nucleic acids in samples. The polypeptide(s)
 CC it encodes could be used as immunoassay reagents and vaccines and to
 CC generate antibodies useful in diagnosis and passive immunotherapy for
 CC HCV infection/non-A, non-B hepatitis.
 CC (Updated on 25-MAR-2003 to correct PR field.)
 XX (Updated on 25-MAR-2003 to correct PI field.)
 SQ Sequence 7310 BP; 1491 A; 2217 C; 2058 G; 1540 T; 4 other;

Alignment Scores:
 Pred. No.: 4,35e-72 Length: 7310
 Score: 901.50 Matches: 175
 Percent Similarity: 90.20% Conservative: 9
 Best Local Similarity: 85.78% Mismatches: 11
 Query Match: 87.35% Indels: 9
 DB: 10 Gaps: 1

US-09-965-594-14 (1-197) x AAN92106 (1-7310)
 QY 3 LysLysGlySerValValIleValGlyArgIleAsn----- 14
 DB 1665 CGCAGGGCGGGAGATACCTCTCGGGCCAGCCCGATGGATGGTCTCCAGGGGTGGAGG 1724
 QY 15 ---LeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGlyCysGlnGlu 33
 DB 1725 TTGCTGGCGCCCATCAGCGCTAGCGCCAGACAGAGGGGCTCTAGGGTGTATATC 1784
 QY 34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 53
 DB 1785 ACCAGCCTAACTGCGCGGACAAAACCAAGTGGAGGTGAGGTCCAGATTGTCTCACT 1844
 QY 54 AlaAlaGlnThrPheLeuAlaThrCysLeuAsnGlyValCysTrpThrValTyrHisGly 73
 DB 1845 GCTGCCCAAACTTCTCGGCAACGTGTCATCAATGGGGTGTGCTGACTGTCTACACGGG 1904
 QY 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 93
 DB 1905 GCCGGAACGAGGACCATCGCTCACCAGGGTCTGTCTATCCAGATGTATACCAATGTA 1964
 QY 94 AspLysAspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThr 113
 DB 1965 GACCAAGACCTTGTGGCTGGCGCGCTCCGCAAGGTAGCCGCTCATTCACACCTGCACT 2024
 QY 114 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 133
 DB 2025 TGGCGCTCTCGGACCTTACTGTGTACAGAGGACGCCGATGTATCCGTCGCGCGG 2084
 QY 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 153
 DB 2085 CGGGGTATAGCAGGGGAGCGCTGTGTGCGCGCGCCATTTCTACTTTGAAAGGCTCC 2144
 QY 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 173
 DB 2145 TCGGGGGTCTCGCTGTGTGCGCGCGGGACGCCGCTAGGCATATTAGGGCGCGGTG 2204
 QY 174 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 193
 DB 2205 TGCACCGGTGGATGGCTAAGCGGTGACCTTTATCCCTGTGGAGAACCTTAGACACAACC 2264
 QY 194 MetArgSerPro 197
 DB 2265 ATGAGTCCCGG 2276
 RESULT 15
 ID AAN90336
 XX AAN90336 standard; DNA: 7310 BP.
 AC AAN90336;
 XX 25-MAR-2003 (updated)

DT 19-JUL-2001 (updated)
 DT 01-NOV-1989 (first entry)
 DE Composite hepatitis C virus (HCV) cDNA.
 XX Hepatitis C virus; cDNA; clone 15e; clone k9-1; probe; vaccine; ds.
 XX Pan troglodytes.
 PN GB2212511-A.
 XX 26-JUL-1989.
 XX 18-NOV-1988; 88GB-0027024.
 XX 18-NOV-1987; 87US-0122714.
 PR 30-DEC-1987; 87US-0139886.
 PR 26-FEB-1988; 88US-0161072.
 PR 26-OCT-1988; 88US-0263584.
 XX (CHIR) CHIRON CORPORATION.
 XX Houghton M, Choo QL, Kuo G;
 XX WPI; 1989-215054/30.
 DR P-PSDB; AAP90288.
 XX Hepatitis C virus gene - used for prodn. of polynucleotide probes,
 PT polypeptide(s) and antibodies for diagnosis, prevention and treatment
 PT of infection.
 XX Disclosure; fig 47; 235pp; English.
 CC The sequence shows a composite hepatitis C virus (HCV) cDNA, derived by
 CC aligning clones k9-1 through 15e in 5'-3' direction. The cDNA
 CC encodes antigens which react with antibodies in patients with non-A
 CC non-B hepatitis (NANBH). The cDNA can be used to design probes, or to
 CC synthesise polypeptides, which are used to diagnose HCV-induced NANBH,
 CC to raise antibodies for immunoassay or treatment, or to produce
 CC vaccines. See also AAP90288, and AAN90303-35.
 CC (N.B. This record was resubmitted to correct errors in the sequence.)
 CC (Updated on 25-MAR-2003 to correct PR field.)
 SQ Sequence 7310 BP; 1495 A; 2218 C; 2058 G; 1539 T; 0 other;
 Alignment Scores:
 Pred. No.: 4,35e-72 Length: 7310
 Score: 901.50 Matches: 175
 Percent Similarity: 90.20% Conservative: 9
 Best Local Similarity: 85.78% Mismatches: 11
 Query Match: 87.35% Indels: 9
 DB: 10 Gaps: 1
 US-09-965-594-14 (1-197) x AAN90336 (1-7310)
 QY 3 LysLysGlySerValValIleValGlyArgIleAsn----- 14
 DB 1665 CGCAGGGCGGGAGATACCTCTCGGGCCAGCCCGATGGATGGTCTCCAGGGGTGGAGG 1724
 QY 15 ---LeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGlyCysGlnGlu 33
 DB 1725 TTGCTGGCGCCCATCAGCGCTAGCGCCAGACAGAGGGGCTCTAGGGTGTATATC 1784
 QY 34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 53
 DB 1785 ACCAGCCTAACTGCGCGGACAAAACCAAGTGGAGGTGAGGTCCAGATTGTGTCACT 1844
 QY 54 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGly 73
 DB 1845 GCTGCCCAAACTTCTCGGCAACGTGTCATCAATGGGGTGTGCTGACTGTCTACACGGG 1904
 QY 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 93
 DB 1905 GCGGGAACGAGGACCATCGCTCACCAGGGTCTGTCTATCCAGATGTATACCAATGTA 1964
 QY 94 AspLysAspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThr 113
 DB 1965 GACCAAGACCTTGTGGCTGGCGCGCTCCGCAAGGTAGCCGCTCATTCACACCTGCACT 2024
 QY 114 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 133
 DB 2025 TGGCGCTCTCGGACCTTACTGTGTACAGAGGACGCCGATGTATCCGTCGCGCGG 2084
 QY 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 153
 DB 2085 CGGGGTATAGCAGGGGAGCGCTGTGTGCGCGCGCCATTTCTACTTTGAAAGGCTCC 2144
 QY 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 173
 DB 2145 TCGGGGGTCTCGCTGTGTGCGCGCGGGACGCCGCTAGGCATATTAGGGCGCGGTG 2204
 QY 174 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 193
 DB 2205 TGCACCGGTGGATGGCTAAGCGGTGACCTTTATCCCTGTGGAGAACCTTAGACACAACC 2264
 QY 194 MetArgSerPro 197
 DB 2265 ATGAGTCCCGG 2276
 RESULT 15
 ID AAN90336
 XX AAN90336 standard; DNA: 7310 BP.
 AC AAN90336;
 XX 25-MAR-2003 (updated)

Db	1905	GCCGGAACGAGGACCATCGCGGTCAACCAAGGTCCTGTCTCATCCAGATGTATACCAATGTA	1964
Qy	94	AspLysAspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThr	113
Db	1965	:	
Qy	114	CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg	133
Db	2025		
Qy	134	ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer	153
Db	2085		
Qy	154	SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValAlaGlyIlePheArgAlaAlaVal	173
Db	2145		
Qy	174	CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr	193
Db	2205		
Qy	194	MetArgSerPro	197
Db	2265	ATGAGGTCCCGG	2276

Search completed: August 30, 2003, 19:47:57
Job time : 192.939 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 30, 2003, 19:20:43 ; Search time 1910.31 Seconds
(without alignments)
2506.388 Million cell updates/sec

Title: US-09-965-594-14
Perfect score: 1032
Sequence: 1 MKKGGVIVGRINLSGDTA.....VAKAVDFIPVESLETTMRSP 197

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+_p2n.model -DEV=xl
-O=/cgn2.1/USPTO_spool/US09965594/runat_29082003.151919.28322/app_query.fasta_1.2872
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09965594 -CGN_1_12630.@runat_29082003.151919.28322 -NCPU=6 -ICPU=3
-NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_hic.*
9: gb_est1.*
10: gb_est2.*
11: gb_hic.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: em_gss_hum.*
18: em_gss_inv.*
19: em_gss_pln.*
20: em_gss_vrt.*
21: em_gss_fun.*
22: em_gss_nam.*
23: em_gss_mus.*
24: em_gss_pro.*
25: em_gss_rod.*
26: em_gss_phg.*
27: em_gss_vrl.*
28: gb_gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
C 1	106	10.3	984	10	BF304699	BF304699	601888252
C 2	103.5	10.0	1199	13	BQ892487	BQ892487	AGENCOURT
C 3	99	9.6	615	12	BQ001625	BQ001625	BQ001625
C 4	99	9.6	643	12	BQ024121	BQ024121	BQ024121
C 5	99	9.6	754	12	BQ016176	BQ016176	BQ016176
C 6	98.5	9.5	961	10	BF203316	BF203316	601865914
C 7	98.5	9.5	1403	13	BQ926101	BQ926101	AGENCOURT
C 8	98	9.5	1141	11	AK080545	AK080545	Mus muscu
C 9	97.5	9.4	779	10	BF631437	BF631437	HVSMB001
C 10	96	9.3	701	10	BF863244	BF863244	963042C02
C 11	96	9.3	846	10	BF182274	BF182274	601804028
C 12	95.5	9.3	901	10	BF307233	BF307233	601891502
C 13	95	9.2	407	9	AW785806	AW785806	117260 MA
C 14	95	9.2	931	13	BQ878887	BQ878887	AGENCOURT
C 15	94.5	9.2	641	9	AU127824	AU127824	AU127824
C 16	94.5	9.2	844	12	B1198486	B1198486	602760491
C 17	94.5	9.2	938	13	BQ894657	BQ894657	AGENCOURT
C 18	94	9.1	1283	13	BQ709745	BQ709745	AGENCOURT
C 19	93.5	9.1	701	14	CD262790	CD262790	pSMA0194E
C 20	93.5	9.1	832	10	BG387051	BG387051	602454749
C 21	93.5	9.1	905	13	BU542842	BU542842	AGENCOURT
C 22	93.5	9.1	993	9	AL555424	AL555424	AL555424
C 23	93.5	9.1	1291	10	BE622016	BE622016	601440668
C 24	93	9.0	960	13	BQ955406	BQ955406	AGENCOURT
C 25	93	9.0	1146	12	BM915803	BM915803	AGENCOURT
C 26	93	9.0	1384	29	CC221189	CC221189	CH261-183
C 27	93	9.0	1637	11	AK038857	AK038857	Mus muscu
C 28	93	9.0	1702	11	AK081278	AK081278	Mus muscu
C 29	92.5	9.0	556	14	CB216999	CB216999	NISC_nq11
C 30	92.5	9.0	582	14	CB286751	CB286751	CMD45_C08
C 31	92.5	9.0	715	9	AU125614	AU125614	AU125614
C 32	92.5	9.0	846	13	BU540812	BU540812	AGENCOURT
C 33	92.5	9.0	866	13	BX451426	BX451426	BX451426
C 34	92.5	9.0	881	14	CD105862	CD105862	AGENCOURT
C 35	92.5	9.0	929	13	BQ672290	BQ672290	AGENCOURT
C 36	92.5	9.0	947	13	BU556872	BU556872	AGENCOURT
C 37	92.5	9.0	958	10	BG420860	BG420860	602452062
C 38	92.5	9.0	979	13	BQ673186	BQ673186	AGENCOURT
C 39	92.5	9.0	1008	12	BF755608	BF755608	603027112
C 40	92	8.9	871	10	BG178418	BG178418	602330206
C 41	92	8.9	898	10	BG385514	BG385514	602453808
C 42	92	8.9	963	10	BF794182	BF794182	602255566
C 43	92	8.9	1001	13	BQ928211	BQ928211	AGENCOURT
C 44	92	8.9	1640	10	BF180599	BF180599	601808704
C 45	91.5	8.9	422	14	CB763743	CB763743	AMGNNUC:S

ALIGNMENTS

RESULT 1
BF304699/c
LOCUS 601888252F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4122276 5',
DEFINITION mRNA sequence.
ACCESSION BF304699
VERSION BF304699.1 GI:11251586
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 984)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: LCM1005 row: 9 column: 13
 High quality sequence stop: 646.
 Location/Qualifiers

FEATURES
 source
 1. 984
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone_image="IMAGE:412276"
 /tissue_type="rhabdomyosarcoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC.17"
 /note="Organ: muscle; Vector: pOTB7; Site.1: EcoRI;
 Site.2: XhoI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCAGG(G). Size-selected >500bp
 for average insert size 1.8kb. Library constructed by
 Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 133 a 329 c 351 g 171 t
 ORIGIN

Alignment Scores:
 Pred. No.: 4.67 Length: 984
 Score: 106.00 Matches: 33
 Percent Similarity: 45.24% Conservative: 5
 Best Local Similarity: 39.29% Mismatches: 24
 Query Match: 10.27% Indels: 22
 DB: 10 Gaps: 5

US-09-965-594-14 (1-197) x BF304699 (1-984)

Qy 100 TrpProAlaProGlnGlySerArgSerLeuThr---ProCysThrCysGlySerSerAsp 118
 Db 646 TGGCCCATGTCACGGCATTCGGTGGGAGAGAGACCGGTACCTGC-----599
 Qy 119 LeuTyrLeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArg 138
 Db 598 -----ACCAGGACGACGACCAACATACATCAAGAGACGTGCT---TCCCGC 554
 Qy 139 GlySerLeuLeuSerProArgPro-----IleSerTyrLeuLysGlySer 153
 Db 553 GGGCGCCTCTGTGGGAGAACCTCGATGCTGTCACAGCTCGCGCTGCTACTGGAAGT 494
 Qy 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 173
 Db 493 CGGCACGCTCGGTACGTGCAGC-----TTCCAGCCCGCGGG 455
 Qy 174 CysThrArgGly 177
 Db 454 TGCGCCGAGGA 443

RESULT 2
 BQ892487
 LOCUS
 DEFINITION BQ892487 1199 bp mRNA linear EST 16-AUG-2002
 IMAGE:6192708 5', mRNA sequence.
 ACCESSION BQ892487
 VERSION BQ892487.1 GI:22284501
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1199)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs@mail.nih.gov
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM13595 row: c column: 13
 High quality sequence start: 57
 High quality sequence stop: 394.
 Location/Qualifiers

FEATURES
 source
 1. 1199
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone_image="IMAGE:6192708"
 /sex="male"
 /tissue_type="sympathetic trunk"
 /dev_stage="adult, 16 yr"
 /lab_host="DH10B"
 /clone_lib="Lupski sympathetic trunk"
 /note="Vector: pCMV-SPORT6 (Life Technologies); Site.1:
 NotI; Site.2: SalI; cDNA made by oligo-dT priming.
 Directionally cloned using the following adaptors:
 5'-TGACCCACGCGTCG-3' and
 5'-GACTATGTCATGATCGAGGCGCGCT(15)-3'. Size selected >
 1 kb for average insert length 1.9 kb. This is a primary
 library, non-amplified. Library constructed by Life
 Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
 College of Medicine); available through Life
 Technologies."

BASE COUNT 255 a 362 c 343 g 211 t 28 others
 ORIGIN

Alignment Scores:
 Pred. No.: 10.4 Length: 1199
 Score: 103.50 Matches: 41
 Percent Similarity: 37.42% Conservative: 17
 Best Local Similarity: 26.45% Mismatches: 53
 Query Match: 10.03% Indels: 44
 DB: 13 Gaps: 6

US-09-965-594-14 (1-197) x BQ892487 (1-1199)

Qy 68 TrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLysGlyProValIle 87
 Db 484 TGGGATCCATTTTATAAAGGGTGCTCTGTATATCGGCCGCCACGGCCCGCTGATA 543
 Qy 88 GlnMetTyrThrAsnValAspLysAspLeuValGlyTyrProAlaProGlnGlySerArg 107
 Db 544 CTTCCATTATCCACATGTGCAGTGACTTT-----573
 Qy 108 SerLeuThrProCysThr-----CysGlySerSerAsp 118
 Db 574 ---TGTGCTGCTGCACACACCCCATGCGGATGTTGGGCTTATGTGGAACGGCGAG 630
 Qy 119 LeuTyr-LeuValThr-----ArgHisAlaAspValIleProValArg----- 132
 Db 631 CGGTTTCATTGGCCACTCCCTCTCTATAAAACACGCCAACGTCGTTCATGGGCGGGCT 690
 Qy 133 -----ArgArgGlyAspSerArgGlySerLeuLeu-- 142
 Db 691 GGGTGTGTCAGCGCAAGCGGGGTGGGGCATGTTAGGACTCGGGGGCGGATTCCTGT 750
 Qy 143 -----SerProArgProIleSerTyrLeuLys-----GlySerSerG1 155

```

Db      751 AAAACCCACCCTCGGCCACCGATGCGCTAAGCTCCCTTTACAGCCACGCCCGCGG 810
      155 yGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysTh 175
      811 CCCCCTTACATCTCTACCTCGCGCGCGCGGGGAGACAGTGGCGCATACGGGC 870
      175 rArgGlyValAlaLysAlaValAspPheIleProValGluSer 189
      871 TCAGGCGCTTTTAAAGCCCGCGGCTTCGCGCGCGGCGGAAGCA 913

RESULT 3
BJ001625/c
LOCUS      BJ001625 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA025C02 5',
DEFINITION mRNA sequence.
ACCESSION BJ001625
VERSION    BJ001625.1 GI:17364516
KEYWORDS   Oryzias latipes (Japanese medaka)
SOURCE     Oryzias latipes
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
            Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE  1 (bases 1 to 615)
AUTHORS   Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
TITLE     Medaka EST Project in Takeda's lab
JOURNAL   Unpublished
COMMENT   Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
            Location/Qualifiers
            1..615
               /organism="Oryzias latipes"
               /mol_type="mRNA"
               /strain="Hd-rR"
               /db_xref="taxon:8090"
               /clone="MF01SSA025C02"
               /sex="mixture of female and male"
               /tissue_type="whole embryo"
               /dev_stage="segmentation stage 20 - 25"
               /clone_lib="MF01SSA cDNA"
BASE COUNT 140 a 156 c 165 g 144 t
ORIGIN

Alignment Scores:
Pred. No.: 12 Length: 615
Score: 99.00 Matches: 42
Percent Similarity: 33.77% Conservative: 9
Best Local Similarity: 27.81% Mismatches: 50
Query Match: 9.59% Indels: 50
DB: 12 Gaps: 7

US-09-965-594-14 (1-197) x BJ001625 (1-615)

Qy      41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAla 60
      511 AAAAATGACGTAGAACCAAAAGACACAGATCCACACACATCTCTGCTTACGGCT 452
      61 ThrCysIleAsnGlyValCysTrpThrValTyHisGlyAlaGlyThrArgThrIleAla 80
      451 -----TGTGGAGAACCTATCACAGTTCCTGCTTTAGACGACGGCA 410
      81 SerProLys-----GlyProValIleGlnMetTyThrAsnValAspLys 95
      409 GCTCTGCGCGCGGAGGAGCTCTCGGCCAGTTGTG----- 374
      96 AspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrPro----- 111

```

```

Db      373 -----ACTCGTGGAGAGCAAGAGCGTCACCCCGAGCTGTAGG 335
      112 -----CysThrCysGlySerSerAspLeuTyTrLeuValThrArg----- 124
      334 CTGCAGGATCGGATGGGCTCTGCT-----TTGGTTCTCTGCTCTCTCTGGATCA 284
      125 -----HisAlaAspValIleProValArgArgGlyAspSer 137
      283 TCTTCTCACCTGACCTTCCACATCCAGGTGTGCGCAGCGCTGTCTGCAGCGGTGATGG 224
      138 ArgGlySerLeuLeuSerProArg-----ProIleSerTyTrLeuLysGlySerSer 154
      223 AGAGCCCGGACAGCAGCAGTCGGGGTGAATCTCTGCAGGACGCTCTTCACGCGGATCA 164
      155 GlyGlyProLeuLeuCysProAlaGlyHisAla 165
      163 GGAGGACCGACTCGCTGCAGAGCCCTCTGCTGCA 131

RESULT 4
BJ024121
LOCUS      BJ024121 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA143D12 3',
DEFINITION mRNA sequence.
ACCESSION BJ024121
VERSION    BJ024121.1 GI:17377389
KEYWORDS   Oryzias latipes (Japanese medaka)
SOURCE     Oryzias latipes
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
            Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE  1 (bases 1 to 643)
AUTHORS   Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
TITLE     Medaka EST Project in Takeda's lab
JOURNAL   Unpublished
COMMENT   Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
            Location/Qualifiers
            1..643
               /organism="Oryzias latipes"
               /mol_type="mRNA"
               /strain="Hd-rR"
               /db_xref="taxon:8090"
               /clone="MF01SSA143D12"
               /sex="mixture of female and male"
               /tissue_type="whole embryo"
               /dev_stage="segmentation stage 20 - 25"
               /clone_lib="MF01SSA cDNA"
BASE COUNT 171 a 148 c 148 g 176 t
ORIGIN

Alignment Scores:
Pred. No.: 12.7 Length: 643
Score: 99.00 Matches: 42
Percent Similarity: 33.77% Conservative: 9
Best Local Similarity: 27.81% Mismatches: 50
Query Match: 9.59% Indels: 50
DB: 12 Gaps: 7

US-09-965-594-14 (1-197) x BJ024121 (1-643)

Qy      41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAla 60
      242 AAAAATGACGTAGAACCAAAAGACACAGATCCACACACATGTTCTGCTTACGGCT 301
      61 ThrCysIleAsnGlyValCysTrpThrValTyHisGlyAlaGlyThrArgThrIleAla 80

```

```

Db 302 -----TCTTGAGAACCTATCACAGTTCTCTGCTTTAGACGACGGCA 343
Qy 81 SerProLys -----GlyProValIleGlnMetTyrThrAsnValAspLys 95
Db 344 GCTCTGGCGGGCGAGAGCTCTCGGCCAGTTGTG----- 379
Qy 96 AspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrPro----- 111
Db 380 -----ACTCGTGGAGGAGGAGGCGTCCACCCGGAGCTGTAGG 418
Qy 112 -----CysThrCysGlySerAspLeuTyrLeuValThrArg----- 124
Db 419 CTGACGGATCGGATGGCTCTGCT-----TTGGTTCTCTCTCTCTGGATCA 469
Qy 125 -----HisAlaAspValIleProValArgArgGlyAspSer 137
Db 470 TCTTCTCCTCCTGACCTCCACATCCAGGTGTCGCCAGCGCTGTGACGGGTGATGG 529
Qy 138 ArgGlySerLeuLeuSerProArg-----ProIleSerTyrLeuLysGlySerSer 154
Db 530 AGAGCGCGACAGCAGCAGTCTCGGGGTGAATCTCTGCAGGACGCTCTTCACGGCGGATCA 589
Qy 155 GlyGlyProLeuLeuCysProAlaGlyHisAla 165
Db 590 GGAGGACCGACTCGCTGCAGAGCCTCTGCTGCA 622

RESULT 5
BJ016176 754 bp mRNA linear EST 05-DEC-2001
LOCUS BJ016176 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA025C02 3',
DEFINITION mRNA sequence.
ACCESSION BJ016176
VERSION BJ016176.1 GI:17376695
KEYWORDS EST.
SOURCE Oryzias latipes (Japanese medaka)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percormorpha; Atherinomorpha;
Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE 1 (bases 1 to 754)
AUTHORS Kohara,Y., Shin-I,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
TITLE Medaka EST Project in Takeda's lab
JOURNAL Unpublished
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES
Location/Qualifiers
1..754
/organism="Oryzias latipes"
/mol_type="mRNA"
/strain="Hd-r"
/db_xref="taxon:8090"
/clone="MF01SSA025C02"
/sex="mixture of female and male"
/tissue_type="whole embryo"
/dev_stage="segmentation stage 20 - 25"
/clone_lib="MF01SSA cDNA"

BASE COUNT 194 a 181 c 181 g 198 t
ORIGIN

Alignment Scores:
Pred. No.: 15.6 Length: 754
Score: 99.00 Matches: 42
Percent Similarity: 33.77% Conservative: 9
Best Local Similarity: 27.81% Mismatches: 50
Query Match: 9.59% Indels: 50
DB: 12 Gaps: 7

```

```

US-09-965-594-14 (1-197) x BJ016176 (1-754)
Qy 41 LysAsnGlnValGluGlyClnValGlnIleValSerThrAlaAlaGlnThrPheLeuAla 60
Db 242 AAAATATGAGCTGTAGAACCAACACACAGATCCACACACATGTTCTGTTCTACGGGCT 301
Qy 61 ThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
Db 302 -----TGTTGGAGAACCTATCACAGTTCTCTGCTTTAGACGACGGCA 343
Qy 81 SerProLys -----GlyProValIleGlnMetTyrThrAsnValAspLys 95
Db 344 GCTCTGGCGGGCGAGGAGCTCTCGGCCAGTTGTG----- 379
Qy 96 AspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrPro----- 111
Db 380 -----ACTCGTGGAGGAGGAGGCGTCCACCCGGAGCTGTAGG 418
Qy 112 -----CysThrCysGlySerAspLeuTyrLeuValThrArg----- 124
Db 419 CTGACGGATCGGATGGCTCTGCT-----TTGGTTCTCTCTCTCTGGATCA 469
Qy 125 -----HisAlaAspValIleProValArgArgGlyAspSer 137
Db 470 TCTTCTCCTCCTGACCTCCACATCCAGGTGTCGCCAGCGCTGTGACGGGTGATGG 529
Qy 138 ArgGlySerLeuLeuSerProArg-----ProIleSerTyrLeuLysGlySerSer 154
Db 530 AGAGCGCGACAGCAGCAGTCTCGGGGTGAATCTCTGCAGGACGCTTCACGGCGGATCA 589
Qy 155 GlyGlyProLeuLeuCysProAlaGlyHisAla 165
Db 590 GGAGGACCGACTCGCTGCAGAGCCTCTGCTGCA 622

RESULT 6
BF203316 961 bp mRNA linear EST 06-NOV-2000
LOCUS BF203316 NTH_MGC_17 Homo sapiens cDNA clone IMAGE:4098578 5',
DEFINITION mRNA sequence.
ACCESSION BF203316
VERSION BF203316.1 GI:11096902
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 961)
AUTHORS NTH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-femail.nih.gov
Tissue Procurement: ATCC
cDNA Library preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: L1CM965 row: 1 column: 03
High quality sequence stop: 637.
FEATURES
Location/Qualifiers
1..961
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4098578"
/tissue_type="rhabdomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NTH_MGC_17"
/notes="Organ: muscle; Vector: pOTB7; Site:1: EcoRI;
Site:2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the

```


PUBMED
REFERENCE

10349636

2

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* 10 (10), 1617-1630 (2000)

20499374

PUBMED

11042159

PUBMED
REFERENCE

3

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsu, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)

20530913

PUBMED

11076861

PUBMED
REFERENCE

4

Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischnann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schriml, L. M., Staubli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilmshing, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S., and Hayashizaki, Y. Functional annotation of a full-length mouse cDNA collection. *Nature* 409 (6821), 685-690 (2001)

21085660

PUBMED

11217851

PUBMED
REFERENCE

5

The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase 1 & 2 Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* 420, 563-573 (2002)

6 (bases 1 to 1141)

PUBMED
REFERENCE

6

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanazaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kuribara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y. Direct Submission

Submitted (16-APR-2002)

PUBMED
REFERENCE

7

Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL: http://genome.gsc.riken.go.jp/ URL: http://fantom.gsc.riken.go.jp/.

FEATURES

source

1. .1141
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM,DB:A730082L10"
/db_xref="taxon:10090"
/clone="A730082L10"
/tissue_type="cerebellum"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="7 days neonate"
<1. .587
/note="unnamed protein product: putative weakly similar to zinc finger protein (fragment) [Mus musculus] (PIR|I48722, evidence: FASTY, 50.7%ID, 57.6%length, match=601)"
/codon_start=3
/protein_id="BAC37940.1"
/db_xref="GI:26348601"
/translation="DSCLPAASGRSLITPRDGFELKELSAARAVGPGSPVAFQVS TRVQAQAGQQQRVGRACRSEGLSKSRPRQRHVPVPGHYIGSGRRIPPPAGE AQAGRAPQGPVPHPPHGTVPVQOAGLLPALAARQVPGVPRGREGPRAPRHS PKVPVTLGFSFGGSGPAPPLAPANGRSVGLAL"

CDS

1118. .1123
polyA_signal
polyA_site
1141
/note="putative"
BASE COUNT 244 a 316 c 353 g 228 t
ORIGIN
Alignment Scores:
Pred. No.: 33 Length: 1141
Score: 98.00 Matches: 46
Percent Similarity: 35.85% Conservative: 11
Best Local Similarity: 28.93% Mismatches: 52
Query Match: 9.50% Indels: 50
DB: 11 Gaps: 9

US-09-965-594-14 (1-197) x AK080545 (1-1141)

Qy 46 GlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAlaThrCysIleAsnGly 65
Db 303 GGAGAGCGCAGCGAGCTGGACGCGCGCAG-----CGCCACATGGAACA 371
Qy 66 ValCysTrpThrValTyHisGlyAlaGlyThrArgThrIleAlaSerProLysGlyPro 85
Db 336 -----CAAGTCCCGCATCCCGCTGGG-----CCTCCTCAAGGA 389
Qy 86 ValIleGlnMetTyThrAsnValAspLysAspLeuValGlyTrpProAlaProGlnGly 105
Db 372 GTGCTG-----CCTCCTCAAGGA 389
Qy 106 SerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyLeuValThrArgHis 125
Db 390 CGCGCTGGCTTCTCCCTGCA-----CTCGCAGCTCGCAA 425
Qy 126 AlaAspValIleProValArg---ArgArgGlyAspSerArgGlySerLeuLeuSerPro 144
Db 426 GTTCCTGTGACCGGTAGGGCGCGGAGGAGCAAGAGGAGCGCAGACACAGCCCC 485
Qy 145 ArgProIle-----SerTyLeuLysGlySerGlyGlyGlyProLeu 158
Db 486 AAGCGGTTCCTACAGCCTTGGGGTTCCTGTTGGGAGGGTGGGCTGCTCCTCCCTC 545
Qy 159 LeuCysProAla----GlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGly 177

Fax: 919 613 8177

Email: chauser@duke.edu.

Location/Qualifiers

FEATURES

source

1. 701

/organism="Chlamydomonas reinhardtii"

/mol_type="mRNA"

/strain="CC-1690 wild type mt+ 21gr"

/db_xref="taxon:3055"

/clone_lib="C. reinhardtii CC-1690, Stress condition I,

normalized, Lambda Zap II"

/note="vector: pBluescript II SK-; Site.1: EcoRI; Site.2:

XhoI; this library, constructed by John Davies and Jeffrey

McDermott, combines cDNAs from CC-1690 cells grown to

mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min,

1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr

, 4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was

purified from each sample, pooled and cDNA synthesized.

The cDNA was directionally cloned into lambda Zap II

(Stratagene) in the EcoRI (5') and XhoI (3') sites.

pBluescript II SK- plasmids were excised from the lambda

ZAP clones by superinfection with ExAssist (Stratagene)

phage. The library was normalized using method 4 described

in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT 173 a 213 c 175 g 140 t

ORIGIN

Alignment Scores:

Pred. No.:	27.6	Length:	701
Score:	96.00	Matches:	32
Percent Similarity:	40.71%	Conservative:	14
Best Local Similarity:	28.32%	Mismatches:	45
Query Match:	9.30%	Indels:	22
DB:	10	Gaps:	4

US-09-965-594-14 (1-197) x BF863244 (1-701)

Qy	71	TyrHisGlyAlaGlyThrArgThrIleAlaSerProLys-----GlyProVal	86
Db	171	CACCACCATACCTTCCTCAGCTGCTCAGCAACAAATATGCCCCATACGGCCACTA	230
Qy	87	IleGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrProAlaProGlnGlySer	106
Db	231	ACAAAGTTACATACAGG-----AAGACACAGCGCGCTTGCCACACCCCTTGAGCGG	284
Qy	107	ArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeuValThrArgHisAla	126
Db	285	AGAAGCCGACGCGCTGCTGCGGTCATCGCATCTATGCAATCTCCCGCTATCAG	344
Qy	127	AspValIle-----ProValArgArgArgGlyAspSerArg-----	138
Db	345	GAGATCATTTGATGTGCTTGTAGTACCCCAAGAGAGCGCTGGGAGTGGGCATTATAA	404
Qy	139	-----GlySerLeuLeuSerProArgProIleSerTyrLeu	150
Db	405	GAAGGGGACGGAAATTCGTTTGGGAAAGAGTACGCGCCCAAGGCTTGACCAAGTCTA	464
Qy	151	LysGlySerGlyGlyProLeuLeuCysProAlaGly	163
Db	465	CTCCAAGGACGAAATGGGAGCGCTTTCGGGCTGTGGCGGT	503

RESULT 11

BF182274/c

LOCUS 601804028F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:4035102 5',
 mRNA sequence. 846 bp linear EST 31-OCT-2000

DEFINITION BF182274.1 GI:11060416

ACCESSION BF182274

VERSION BF182274.1

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 846)

AUTHORS

TITLE

JOURNAL

COMMENT

NIH-MGC http://mgc.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgabps@email.nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys

cDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM9308 row: g column: 07

High quality sequence stop: 696.

FEATURES

source

1. 846

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="IMAGE:4035102"

/tissue_type="tumor, gross tissue"

/dev_stage="7 months"

/lab_host="DH10B"

/clone_lib="NCI_CGAP_Mam5"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site.1: SalI;

Site.2: NotI; Cloned unidirectionally. Primer: Oligo dT.

Library constructed by Life Technologies. Investigators

providing samples: Lothar Hennighausen/Robin Humphreys,

NIH

BASE COUNT 176 a 218 c 241 g 210 t

ORIGIN

1 others

Alignment Scores:

Pred. No.:	35.1	Length:	846
Score:	96.00	Matches:	46
Percent Similarity:	47.20%	Conservative:	13
Best Local Similarity:	36.80%	Mismatches:	36
Query Match:	9.30%	Indels:	31
DB:	10	Gaps:	9

US-09-965-594-14 (1-197) x BF182274 (1-846)

Qy	73	GlyAlaGlyThrArg-ThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAs	92
Db	757	GGTTCTGCTACCAGAACACGCTGGATGAAGAAAGGACCA-----CATCCTTC	710
Qy	92	nValAspLysAspLeuValGlyTyr-----ProAlaProGl	104
Db	709	GGTTCTTCAGTCCCAAGTGGCTGGAGAAAGTGAACACAGAGAGAGGACGCTCCTCA	650
Qy	104	nGlySerArg-----SerLeuThrProCysThrCysGlySerSerAspLeuVal	122
Db	649	GTCCCCACGGTTAGTAGTCTAGACAGTGTCTGCTGGA-----	610
Qy	122	lThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu	142
Db	609	-ACTAGACACACCT--GTAATCCAGGAGGAACGCTGGAGGAACAGAGGACTCC--CT	556
Qy	142	uSerProArgProIleSerTyrLeuLysGlySerGlyGlyProLeu---LeuCysPr	161
Db	555	GACCCCACTCC---TCCCTCTAGCGGCACCTCTCTGCGCCCACTCCCTCTGTCC	499
Qy	161	oAlaGlyHis---AlaValGlyIlePheArg-----AlaAlaValCysThrAr	176
Db	498	TAGTGGGCACCTCTCCCGAGGACGACGACTGTACTCCCTTTGGCCCTCTGCACCT	439
Qy	176	gGlyValAlaLys	180
Db	438	TGGGATGACTGAG	426
RESULT 12			
BF307233			

```

LOCUS       BF307233               901 bp      mRNA      linear      EST 21-NOV-2000
DEFINITION   601891502F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4137145 5',
            mRNA sequence.
ACCESSION   BF307233
VERSION     BF307233.1   GI:11254342
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 901)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-re@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCW1044 row: C column: 02
            High quality sequence start: 6
            High quality sequence stop: 684.
            Location/Qualifiers
                1..901
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:4137145"
                /issue_type="rhabdomyosarcoma"
                /lab_host="DH10B (phage-resistant)"
                /clone_lib="NIH_MGC_17"
                /note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI;
                Site_2: XhoI; cDNA made by oligo-dT priming.
                Directionally cloned into EcoRI/XhoI sites using the
                following 5' adaptor: GGCACGAG(G). Size-selected >500bp
                for average insert size 1.8kb. Library constructed by
                Ling Hong in the laboratory of Gerald M. Rubin (University
                of California, Berkeley) using ZAP-cDNA synthesis kit
                (Stratagene) and Superscript II RT (Life Technologies)."
```

```

BASE COUNT   144 a   267 c   329 g   161 t
ORIGIN
1..901
Alignment Scores:
Pred. No.:      42.5      Length:      901
Score:          95.50     Matches:     37
Percent Similarity: 38.46% Conservative:  8
Best Local Similarity: 31.62% Mismatches:   28
Query Match:     9.25%   Indels:      45
DB:              10     Gaps:       5

US-09-965-594-14 (1-197) x BF307233 (1-901)
QY      66 ValCysTrpThr-ValTyrHis-----GlyAlaGlyThrArgThrIleAl 80
         ::::::::::: ||| |||
Db      620 ATGTGTTGGAGCGTCCCGCACGGCCATCTGAGGCGGGTCCGGCACACACGAGCTGG 679
QY      80 aSerProLysGlyProValIleGlnMetTyrThrAsnValAspLysAspLeuValGlyTr 100
         ||| |||
Db      680 TGGGCGAGGAGGTGGAGTGTGTGCA-----GTGTGAGGATG 715
QY      100 pProAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTy 120
         ||||||||| |||
Db      716 GCCCGCCCATCCGGG-----
QY      120 rLeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySe 140
         ::::::::::: |||
Db      732 -----GTGAGCCCTCGTCTCAGGCGGCTTGGGGGGGGTTC 766
QY      140 rLeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCy 160
         ||||||| |||

```

```

Db      767 CCTTTTGGGTCTCTGA-----CGGGCATCTCTCCAGGGCGCGCTGGACTG 810
QY      160 sprAlaGlyHisAlaValGly-----IlePheArgAlaAlaValCys 174
         ||||||||| ||| |||
Db      811 TCCGGCGGTTCGCCCGGGCGGCCACACAGGGTCGGCGGGCGCTGTGC 859

RESULT 13
AW785806
LOCUS       117260 MARC 1P1G Sus scrofa cDNA 5', mRNA linear EST 09-JUL-2000
DEFINITION   AW785806
ACCESSION   AW785806
VERSION     AW785806.1   GI:7842582
KEYWORDS    EST.
SOURCE      Sus scrofa (pig)
ORGANISM    Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1 (bases 1 to 407)
AUTHORS     Fahrenkrug,S.C., Smith,T.P.L., Freking,B.A., Cho,J., White,J.,
            Vallet,J., Wise,T., Rohrer,G.A., Perlea,G., Sultana,R., Quackenbush
            J., and Keele,J.W.
            Porcine gene discovery by normalized cDNA-library sequencing and
            EST cluster assembly
            Mamm. Genome 13 (8), 475-478 (2002)
JOURNAL     Mamm. Genome 13 (8), 475-478 (2002)
MEDLINE     22213789
PUBMED      12226715
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smithdemail.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR Primers
            FORWARD: AGGAACAGCTATGACCAI
            BACKWARD: GTTTCCTCCAGTCACGAG
            Plate: 37 row: D column: 16
            Seq primer: ATTTAGGTGACACTATAG.
            Location/Qualifiers
                1..407
                /organism="Sus scrofa"
                /mol_type="mRNA"
                /db_xref="taxon:9823"
                /issue_type="pooled"
                /lab_host="DH10B"
                /clone_lib="MARC 1P1G"
                /note="Vector: PCMV SPOR6; Site_1: NotI; Site_2: SalI;
                Library made from pooled tissue from day 11, 13, 15, 20,
                and 30 embryos."
```

```

BASE COUNT   55 a   131 c   145 g   76 t
ORIGIN
1..407
Alignment Scores:
Pred. No.:      17.2      Length:      407
Score:          95.00     Matches:     41
Percent Similarity: 31.21% Conservative:  3
Best Local Similarity: 29.08% Mismatches:   52
Query Match:     9.21%   Indels:      45
DB:              9     Gaps:       7

US-09-965-594-14 (1-197) x AW785806 (1-407)
QY      43 GlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAlaThrCys 62
         ||| |||
Db      61 CAACCCCGGAGCTCTTCGGCCGATCCCTATCGCTTCGTGGTGTGATGCGAGATGC 120
QY      63 IleAsnGlyValCysTrpThrVal-----TyrHis-Gl 73
         ::::::: |||||
Db      121 GTTTCGACGGCTGTGGGCTTCCGGGGGCTTCGTGCACCGGGCTTCTGGTATTGG 180
QY      73 yAlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVa 93

```



```
FEATURES
source      Location/Qualifiers
1..641
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="NT2RP2002160"
/cell_type="teratocarcinoma"
/cell_line="NT2"
/clone_lib="NT2RP2"
/note="Vector: pME18SFL3: mRNA from NT2 neuronal precursor
cells after 2-weeks retinoic acid (RA) induction"
BASE COUNT      103 a   256 c   183 g   96 t       3 others
ORIGIN

Alignment Scores:
Pred. No.:      34.3      Length:      641
Score:          94.50      Matches:    47
Percent Similarity: 39.10%  Conservative: 14
Best Local Similarity: 30.13%  Mismatches: 56
Query Match:      9.16%     Indels:    39
DB:               9        Gaps:      9

US-09-965-594-14 (1-197) x AUI27824 (1-641)

QY 67 CysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLysGlyProVal 86
Db 154 TGCACGACGACACCCCTACCGCGGGGAGCAGCCGCCACCCACCCACCGCGGCCCTGGC 213
QY 87 IleGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpProAlaProGlnGlySer 106
Db 214 GGGCAATG-----ACATCCTGGCCAGCCCCCTCGCCTGC 249
QY 107 ArgSerLeuThrPro-----CysThrCysGlySer 116
Db 250 CCAGCCCCAGCCCTACCCCGGAGCCCGCCACACAGCTCCTACGTGCACCTGCGGCCGG 309
QY 117 SerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgGlyAsp 136
Db 310 CACGAC-----CCACAAGCCACCCG-----CCACAGCCACCGGACTTC 351
QY 137 SerArgGlySer-----LeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 154
Db 352 AGCCGGTGCTCCACCTGGTTGGCTCAACAGCCGCCCTGTCAAGCGGCATGCGGGCATCC 411
QY 155 GlyGlyPro-Leu---LeuCysProAlaGly-HisAlaValGlyIle-----PheA 170
Db 412 GCGGGCGGACTTCCAGTGCTTCAGCAGCGCGGGCCGTGGGGCTGGCGGCGACCTTC 471
QY 170 rgAlaAlaValCysThrArg-----GlyValAlaLysAlaValAsp---- 183
Db 472 GCACCTTCCTGCTCTCGCCTGCAGGACCTGTACGGCATGTCGGCGCTGGCGGCGCG 531
QY 184 --PheIleProValGluSerLeuGluThrThrMetArgSerPro 197
Db 532 CAACCGTGCCCATGCTCAACCTCAAGGACGAACTGCTGTTTCCA 575
```

Search completed: August 31, 2003, 04:27:28
Job time : 1914.31 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 17:42:58 : Search time 44.6227 Seconds
(without alignments)
700.745 Million cell updates/sec

Title: US-09-965-594-16

Perfect score: 1032

Sequence: 1 MKKKGSWIVGRINLSGDTA.....YAKAVDFIPVESLETTMRSP 197

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03:*

- 1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
- 24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1032	100.0	197	21	AA15222 Hepatitis C virus
2	1015	98.4	197	21	AA15221 Hepatitis C virus
3	1002	97.1	197	21	AA15223 Hepatitis C virus
4	990	95.9	197	21	AA15224 Hepatitis C virus
5	981	95.1	195	21	AA15220 Hepatitis C virus
6	980	95.0	197	21	AA15225 Hepatitis C virus
7	976	94.6	197	21	AA15226 Hepatitis C virus
8	942	91.3	195	21	AA15212 Hepatitis C virus
9	911.5	88.3	665	20	AA124943 HCV NS4A-NS3 compl

10	908.5	88.0	665	20	AA124947	HCV NS4A-NS3 compl
11	907.5	87.9	665	20	AA124941	HCV NS4A-NS3 compl
12	907.5	87.9	665	20	AA124942	HCV NS4A-NS3 compl
13	904.5	87.6	665	20	AA124945	HCV NS4A-NS3 compl
14	904.5	87.6	665	20	AA124946	HCV NS4A-NS3 compl
15	904.5	87.6	665	20	AA124940	HCV NS4A-NS3 compl
16	903.5	87.5	665	20	AA124948	HCV NS4A-NS3 compl
17	903.5	87.5	671	20	AA124948	HCV NS4A-NS3 compl
18	901.5	87.4	216	20	AA124948	HCV NS4A-NS3 compl
19	900.5	87.3	216	20	AA124948	HCV NS4A-NS3 compl
20	900.5	87.3	216	20	AA124948	HCV NS4A-NS3 compl
21	900.5	87.3	665	20	AA124949	HCV NS4A-NS3 compl
22	900.5	87.3	671	20	AA124949	HCV NS4A-NS3 compl
23	900	87.2	215	20	AA124949	HCV NS4A-NS3 compl
24	897.5	87.0	216	20	AA124949	HCV NS4A-NS3 compl
25	897.5	87.0	216	20	AA124949	HCV NS4A-NS3 compl
26	897.5	87.0	216	20	AA124949	HCV NS4A-NS3 compl
27	896.5	86.9	216	20	AA124949	HCV NS4A-NS3 compl
28	894	86.6	215	20	AA124949	HCV NS4A-NS3 compl
29	893.5	86.6	216	20	AA124949	HCV NS4A-NS3 compl
30	893.5	86.6	216	20	AA124949	HCV NS4A-NS3 compl
31	889	86.1	213	20	AA124949	HCV NS4A-NS3 compl
32	889	86.1	631	20	AA124949	HCV NS4A-NS3 compl
33	888.5	86.1	3011	13	AA124949	Hepatitis C virus
34	888.5	86.1	3011	24	AA124949	Amino acid sequenc
35	888.5	86.1	3012	23	AA124949	Hepatitis C virus
36	885.5	85.8	3011	14	AA124949	HCV genomic amino
37	884.5	85.7	687	16	AA124949	pHCV150-encoded se
38	884.5	85.7	1648	16	AA124949	pHCV176-encoded se
39	884.5	85.7	1766	10	AA124949	Sequence encoded i
40	884.5	85.7	1786	10	AA124949	Protein sequence o
41	884.5	85.7	2261	10	AA124949	Peptide encoded by
42	884.5	85.7	2301	10	AA124949	Sequence encoded i
43	884.5	85.7	2436	10	AA124949	Sequence encoded i
44	884.5	85.7	2436	10	AA124949	Peptide encoded by
45	884.5	85.7	2772	21	AA124949	Protein encoded by

ALIGNMENTS

RESULT 1

AA15222
ID AA15222 standard; protein: 197 AA.
XX
AC AA15222;
XX
DT 19-DEC-2000 (first entry)
XX
DE Hepatitis C virus NS4A-NS3 fusion protease #4.
XX
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW liver failure; liver cancer; mutant; mutein.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
PN WO200040707-A1.
XX
PD 13-JUL-2000.
XX
PF 06-JAN-2000; 2000WO-US00345.
XX
PR 08-JAN-1999; 99US-0115271.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI Wittekand M, Weinheimer S, Zhang Y, Goldfarb V;
XX
DR WPI: 2000-465976/40.
XX
DR N-PSDB; AAA73331.
XX
PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1

PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 XX
 XX
 PS Claim 23; Fig 14; 66pp; English.
 CC The present sequence is a mutated version of a fusion protein created
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
 CC proteins are both essential for the replication of the virus, acting to
 CC cleave its replicative proteins from the polyprotein produced from the
 CC HCV genome. Inhibitors of the two proteins should be effective as
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A
 CC fusion proteins which can be used to identify inhibitors of this type, as
 CC well as enabling structural studies of the protease and
 CC protease-inhibitor complexes. This sequence contains the alpha-helix0-1
 CC variant.
 XX
 XX
 SQ Sequence 197 AA;

Query Match 100.0%; Score 1032; DB 21; Length 197;
 Best Local Similarity 100.0%; Pred. No. 6e-101;
 Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKKGSVVIVGRINLSGDTAYAQOTRGECCQETSGTGRDKNQVEGEVQIVSTATQTFLA 60
 DB 1 MKKGSVVIVGRINLSGDTAYAQOTRGECCQETSGTGRDKNQVEGEVQIVSTATQTFLA 60
 QY 61 TCINGVCWTVYHGAGTRTITASPKGPVTOMYTNVDKDLVGVQAPQGSRSLTPTCTCGSSDLY 120
 DB 61 TCINGVCWTVYHGAGTRTITASPKGPVTOMYTNVDKDLVGVQAPQGSRSLTPTCTCGSSDLY 120
 QY 121 LVTRHADVIPVRRGRDGRSLLSPRISYLKSGSGGPLLCPAGHAGVIFRAAVCTRGVAK 180
 DB 121 LVTRHADVIPVRRGRDGRSLLSPRISYLKSGSGGPLLCPAGHAGVIFRAAVCTRGVAK 180
 QY 181 AVDFIPVESLETTMRSP 197
 DB 181 AVDFIPVESLETTMRSP 197

RESULT 2
 AAB15221
 ID AAB15221 standard; protein; 197 AA.

XX AC AAB15221;
 XX DT 19-DEC-2000 (first entry)
 XX DE Hepatitis C virus NS4A-NS3 fusion protease #3.
 XX KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
 XX KW liver failure; liver cancer; mutant; muten.
 XX OS Hepatitis C virus.
 XX OS Synthetic.
 XX PN WO200040707-A1.
 XX PD 13-JUL-2000.
 XX PF 06-JAN-2000; 2000WO-US00345.
 XX PR 08-JAN-1999; 990S-0115271.
 XX PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
 XX WPI: 2000-465976/40.
 XX DR N-PSDB; AAA73330.

PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 XX
 XX
 PS Claim 23; Fig 13; 66pp; English.
 CC The present sequence is a mutated version of a fusion protein created
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
 CC proteins are both essential for the replication of the virus, acting to
 CC cleave its replicative proteins from the polyprotein produced from the
 CC HCV genome. Inhibitors of the two proteins should be effective as
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A
 CC fusion proteins which can be used to identify inhibitors of this type, as
 CC well as enabling structural studies of the protease and
 CC protease-inhibitor complexes. This sequence contains the alpha-helix0-1
 CC variant.
 XX
 XX
 SQ Sequence 197 AA;

Query Match 98.4%; Score 1015; DB 21; Length 197;
 Best Local Similarity 98.5%; Pred. No. 3.8e-99;
 Matches 194; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MKKGSVVIVGRINLSGDTAYAQOTRGECCQETSGTGRDKNQVEGEVQIVSTATQTFLA 60
 DB 1 MKKGSVVIVGRINLSGDTAYAQOTRGECCQETSGTGRDKNQVEGEVQIVSTAQTFLA 60
 QY 61 TCINGVCWTVYHGAGTRTITASPKGPVTOMYTNVDKDLVGVQAPQGSRSLTPTCTCGSSDLY 120
 DB 61 TCINGVCWTVYHGAGTRTITASPKGPVTOMYTNVDKDLVGVQAPQGSRSLTPTCTCGSSDLY 120
 QY 121 LVTRHADVIPVRRGRDGRSLLSPRISYLKSGSGGPLLCPAGHAGVIFRAAVCTRGVAK 180
 DB 121 LVTRHADVIPVRRGRDGRSLLSPRISYLKSGSGGPLLCPAGHAGVIFRAAVCTRGVAK 180
 QY 181 AVDFIPVESLETTMRSP 197
 DB 181 AVDFIPVESLETTMRSP 197

RESULT 3
 AAB15223
 ID AAB15223 standard; protein; 197 AA.
 XX AC AAB15223;
 XX DT 19-DEC-2000 (first entry)
 XX DE Hepatitis C virus NS4A-NS3 fusion protease #5.
 XX KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
 XX KW liver failure; liver cancer; mutant; muten.
 XX OS Hepatitis C virus.
 XX OS Synthetic.
 XX PN WO200040707-A1.
 XX PD 13-JUL-2000.
 XX PF 06-JAN-2000; 2000WO-US00345.
 XX PR 08-JAN-1999; 990S-0115271.
 XX PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
 XX WPI: 2000-465976/40.
 XX DR N-PSDB; AAA73332.

XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
PT amino acid, useful for screening inhibitors that may treat hepatitis C
PT
XX
PS Claim 23; Fig 15; 66pp; English.
XX
CC The present sequence is a mutated version of a fusion protein created
CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
CC proteins are both essential for the replication of the virus, acting to
CC cleave its replicative proteins from the polyprotein produced from the
CC HCV genome. Inhibitors of the two proteins should be effective as
CC antiviral treatments of HCV infection. This is useful as HCV can lead to
CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
CC The present invention concerns a number of NS3 mutants and NS3-NS4A
CC fusion proteins which can be used to identify inhibitors of this type, as
CC well as enabling structural studies of the protease and
CC protease:inhibitor complexes. This sequence contains the alpha-helix0-1
CC variant.
XX
SQ Sequence 197 AA;
Query Match 97.1%; Score 1002; DB 21; Length 197;
Best Local Similarity 98.5%; Pred. No. 9e-98;
Matches 194; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 MKKGSVIVGRINLSGDTAYAAQQTREGGCGQETSGTGRDKNQVEGEVQIVSTATQTFLA 60
DB 1 MKKGSVIVGRINLSGDTAYAAQQTREGGCGQETSGTGRDKNQVEGEVQIVSTATQTFLA 60
QY 61 TCINGVCTVYHGAGTRTIIASPKGPVTQMTYTVNVDKDLVGMQAPQGSRSLSLTPTCTGSSDLY 120
DB 61 TSINGVLTVYHGAGTRTIIASPKGPVTQMTYTVNVDKDLVGMQAPQGSRSLSLTPTCTGSSDLY 120
QY 121 LVTRHADVIPVRRGDSRGLSPRPISYLKSGSGGPLLCPAGHAGVGFRAAVCTRGVAK 180
DB 121 LVTRHADVIPVRRGDSRGLSPRPISYLKSGSGGPLLCPAGHAGVGFRAAVCTRGVAK 180
QY 181 AVDFIPVESLETTMRSP 197
DB 181 AVDFIPVESLETTMRSP 197
RESULT 4
AAB15224
ID AAB15224 standard; protein; 197 AA.
XX
AC AAB15224;
XX
DT 19-DEC-2000 (first entry)
XX
DE Hepatitis C virus NS4A-NS3 fusion protease #6.
XX
XX Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW liver failure; liver cancer; mutant; mutein.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
PN WO200040707-A1.
XX
PD 13-JUL-2000.
XX
PF 06-JAN-2000; 2000WO-US00345.
XX
PR 08-JAN-1999; 99US-0115271.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX WPI; 2000-465976/40.
XX

DR N-PSDB; AAA73333.
XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
PT amino acid, useful for screening inhibitors that may treat hepatitis C
PT
XX
PS Claim 23; Fig 16; 66pp; English.
XX
CC The present sequence is a mutated version of a fusion protein created
CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
CC proteins are both essential for the replication of the virus, acting to
CC cleave its replicative proteins from the polyprotein produced from the
CC HCV genome. Inhibitors of the two proteins should be effective as
CC antiviral treatments of HCV infection. This is useful as HCV can lead to
CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
CC The present invention concerns a number of NS3 mutants and NS3-NS4A
CC fusion proteins which can be used to identify inhibitors of this type, as
CC well as enabling structural studies of the protease and
CC protease:inhibitor complexes. This sequence contains the alpha-helix0-7
CC variant.
XX
SQ Sequence 197 AA;
Query Match 95.9%; Score 990; DB 21; Length 197;
Best Local Similarity 97.0%; Pred. No. 1.7e-96;
Matches 191; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 1 MKKGSVIVGRINLSGDTAYAAQQTREGGCGQETSGTGRDKNQVEGEVQIVSTATQTFLA 60
DB 1 MKKGSVIVGRINLSGDTAYAAQQTREGGCGQETSGTGRDKNQVEGEVQIVSTATQTFLA 60
QY 61 TCINGVCTVYHGAGTRTIIASPKGPVTQMTYTVNVDKDLVGMQAPQGSRSLSLTPTCTGSSDLY 120
DB 61 TSINGVLTVYHGAGTRTIIASPKGPVTQMTYTVNVDKDLVGMQAPQGSRSLSLTPTCTGSSDLY 120
QY 121 LVTRHADVIPVRRGDSRGLSPRPISYLKSGSGGPLLCPAGHAGVGFRAAVCTRGVAK 180
DB 121 LVTRHADVIPVRRGDSRGLSPRPISYLKSGSGGPLLCPAGHAGVGFRAAVCTRGVAK 180
QY 181 AVDFIPVESLETTMRSP 197
DB 181 AVDFIPVESLETTMRSP 197
RESULT 5
AAB15220
ID AAB15220 standard; protein; 195 AA.
XX
AC AAB15220;
XX
DT 19-DEC-2000 (first entry)
XX
DE Hepatitis C virus NS4A-NS3 fusion protease #2.
XX
XX Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW liver failure; liver cancer; mutant; mutein.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
PN WO200040707-A1.
XX
PD 13-JUL-2000.
XX
PF 06-JAN-2000; 2000WO-US00345.
XX
PR 08-JAN-1999; 99US-0115271.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX

DR WPI; 2000-465976/40.
XX N-PSDB; AAA73329.

PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
PT amino acid, useful for screening inhibitors that may treat hepatitis C

XX Claim 23; Fig 12; 66pp; English.

XX The present sequence is a mutated version of a fusion protein created
CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
CC proteins are both essential for the replication of the virus, acting to
CC cleave its replicative proteins from the polyprotein produced from the
CC HCV genome. Inhibitors of the two proteins should be effective as
CC antiviral treatments of HCV infection. This is useful as HCV can lead to
CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
CC The present invention concerns a number of NS3 mutants and NS3-NS4A
CC fusion proteins which can be used to identify inhibitors of this type, as
CC well as enabling structural studies of the protease and
CC protease:inhibitor complexes. This sequence contains the alpha-helix0-1
CC variant.

XX Sequence 195 AA;

Query Match 95.1%; Score 981; DB 21; Length 195;
Best Local Similarity 96.4%; Pred. No. 1.5e-95;
Matches 190; Conservative 1; Mismatches 4; Indels 2; Gaps 1;

OY 1 MKKGSVVIVGRINLSGDTAYAAQOTRGECCOETSQTGRDNQVEGEVQIVSTATQTFLA 60
DB 1 MKKGSVVIVGRIVLNG--AYAAQOTRGECCOETSQTGRDNQVEGEVQIVSTATQTFLA 58
OY 61 TCINGVCWTVYHGAGTRTITASPKGPVTOMYTNVDKDLVGMQAPQGSRSILTPCTCGSSDLY 120
DB 59 TCINGVCWTVYHGAGTRTITASPKGPVQIMYTNVDKDLVGMQAPQGSRSILTPCTCGSSDLY 118
OY 121 LVTRHADVIPVRRRGDSRGSLLSPRISYLGSGGGLLCPAGHAGVIFRAAVCTRGVAK 180
DB 119 LVTRHADVIPVRRRGDSRGSLLSPRISYLGSGGGLLCPAGHAGVIFRAAVCTRGVAK 178
OY 181 AVDFIPVESLETTMRSP 197
DB 179 AVDFIPVESLETTMRSP 195

RESULT 6
AAB15225
ID AAB15225 standard; protein; 197 AA.
XX AAB15225;
AC AAB15225;
DT 19-DEC-2000 (first entry)
XX Hepatitis C virus NS4A-NS3 fusion protease #7.

XX Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW liver failure; liver cancer; mutant; mutain.
XX Hepatitis C virus.
OS Synthetic.
XX WO200040707-A1.
PN 13-JUL-2000.
PD 06-JAN-2000; 2000WO-US00345.
PF 08-JAN-1999; 99US-0115271.
XX (BRIM) BRISTOL-MYERS SQUIBB CO.
PA Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;

XX WPI; 2000-465976/40.
DR N-PSDB; AAA73334.

XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
PT amino acid, useful for screening inhibitors that may treat hepatitis C

XX Claim 23; Fig 17; 66pp; English.

XX The present sequence is a mutated version of a fusion protein created
CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
CC proteins are both essential for the replication of the virus, acting to
CC cleave its replicative proteins from the polyprotein produced from the
CC HCV genome. Inhibitors of the two proteins should be effective as
CC antiviral treatments of HCV infection. This is useful as HCV can lead to
CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
CC The present invention concerns a number of NS3 mutants and NS3-NS4A
CC fusion proteins which can be used to identify inhibitors of this type, as
CC well as enabling structural studies of the protease and
CC protease:inhibitor complexes. This sequence contains the alpha-helix0-7
CC variant.

XX Sequence 197 AA;

Query Match 95.0%; Score 980; DB 21; Length 197;
Best Local Similarity 96.4%; Pred. No. 1.9e-95;
Matches 190; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 MKKGSVVIVGRINLSGDTAYAAQOTRGECCOETSQTGRDNQVEGEVQIVSTATQTFLA 60
DB 1 MKKGSVVIVGRINLSGDTAYAAQOTRGECCOETSQTGRDNQVEGEVQIVSTATQTFLA 60
OY 61 TCINGVCWTVYHGAGTRTITASPKGPVTOMYTNVDKDLVGMQAPQGSRSILTPCTCGSSDLY 120
DB 61 TSINGVLMTVYHGAGTRTITASPKGPVTOMYTNVDKDLVGMQAPQGSRSILTPCTCGSSDLY 120
OY 121 LVTRHADVIPVRRRGDSRGSLLSPRISYLGSGGGLLCPAGHAGVIFRAAVCTRGVAK 180
DB 121 LVTRHADVIPVRRRGDSRGSLLSPRISYLGSGGGLLCPAGHAGVIFRAAVCTRGVAK 180
OY 181 AVDFIPVESLETTMRSP 197
DB 181 AVDFIPVESLETTMRSP 197

RESULT 7
AAB15226
ID AAB15226 standard; protein; 197 AA.
XX AAB15226;
AC AAB15226;
DT 19-DEC-2000 (first entry)
XX Hepatitis C virus NS4A-NS3 fusion protease #8.

XX Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW liver failure; liver cancer; mutant; mutain.
XX Hepatitis C virus.
OS Synthetic.
XX WO200040707-A1.
PN 13-JUL-2000.
PD 06-JAN-2000; 2000WO-US00345.
PF 08-JAN-1999; 99US-0115271.
XX (BRIM) BRISTOL-MYERS SQUIBB CO.

PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;

XX WPI: 2000-465976/40.

DR N-PSDB; AAA73335.

XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1

PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic

PT amino acid, useful for screening inhibitors that may treat hepatitis C

PT -

XX Example 5; Fig 18; 66pp; English.

XX The present sequence is a mutated version of a fusion protein created
CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
CC proteins are both essential for the replication of the virus, acting to
CC cleave its replicative proteins from the polyprotein produced from the
CC HCV genome. Inhibitors of the two proteins should be effective as
CC antiviral treatments of HCV infection. This is useful as HCV can lead to
CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
CC The present invention concerns a number of NS3 mutants and NS3-NS4A
CC fusion proteins which can be used to identify inhibitors of this type, as
CC well as enabling structural studies of the protease and
CC protease:inhibitor complexes. This sequence contains the alpha-helix0
CC wild-type sequence.

XX Sequence 197 AA;

Query Match 94.6%; Score 976; DB 21; Length 197;

Best Local Similarity 95.9%; Pred. No. 5.1e-95;

Matches 189; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 MKKKGSVVIVGRINLSGDTAYAAQOTRGEGCGOETSGTGRDKKNQVEGEVOIVSTATQTFIA 60

Db 1 MKKKGSVVIVGRINLSGDTAYAAQOTRGEGCGOETSGTGRDKKNQVEGEVOIVSTAAQTFLA 60

QY 61 TCINGVCWTVYHGAGTGTIASPKGPVTOMYTNVDKDLVGMWPAQGSRLTPTCTCGSSDLY 120

Db 61 TCINGVCWTVYHGAGTGTIASPKGPVTOMYTNVDKDLVGMWPAQGSRLTPTCTCGSSDLY 120

QY 121 LVTRHADVIPVRRRGDSRGLSPRPISYLGKSSGGPLLCPCAGHAGVIFRAAVCTRGVAK 180

Db 121 LVTRHADVIPVRRRGDSRGLSPRPISYLGKSSGGPLLCPCAGHAGVIFRAAVCTRGVAK 180

QY 181 AVDFIPVESLETMRSP 197

Db 181 AVDFIPVESLETMRSP 197

RESULT 8

AAB15212

ID AAB15212 standard; protein; 195 AA.

XX AAB15212;

XX 19-DEC-2000 (first entry)

XX Hepatitis C virus NS4A-NS3 fusion protease #1.

XX Hepatitis; NS3 protease; viral replication; chronic liver disease;

KW liver failure; liver cancer.

XX Hepatitis C virus.

OS Synthetic.

XX WO2000040707-A1.

XX 13-JUL-2000.

XX 06-JAN-2000; 2000WO-US00345.

XX 08-JAN-1999; 99US-0115271.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

XX

PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;

XX WPI: 2000-465976/40.

DR N-PSDB; AAA73328.

XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
PT amino acid, useful for screening inhibitors that may treat hepatitis C
PT -

XX Example 2; Fig 10; 66pp; English.

XX The present sequence is a fusion protein created using the Hepatitis C
CC virus (HCV) NS3 and NS4A protease enzymes. These proteins are both
CC essential for the replication of the virus, acting to cleave its
CC replicative proteins from the polyprotein produced from the HCV genome.
CC Inhibitors of the two proteins should be effective as antiviral
CC treatments of HCV infection. This is useful as HCV can lead to chronic
CC liver disease such as cirrhosis, liver failure and liver cancer. The
CC present invention concerns a number of NS3 mutants and NS3-NS4A fusion
CC proteins which can be used to identify inhibitors of this type, as well
CC as enabling structural studies of the protease and protease:inhibitor
CC complexes.

XX Sequence 195 AA;

Query Match 91.3%; Score 942; DB 21; Length 195;

Best Local Similarity 93.9%; Pred. No. 2e-91;

Matches 185; Conservative 1; Mismatches 9; Indels 2; Gaps 1;

QY 1 MKKKGSVVIVGRINLSGDTAYAAQOTRGEGCGOETSGTGRDKKNQVEGEVOIVSTATQTFIA 60

Db 1 MKKKGSVVIVGRINLSGDTAYAAQOTRGEGCGOETSGTGRDKKNQVEGEVOIVSTAAQTFLA 58

QY 61 TCINGVCWTVYHGAGTGTIASPKGPVTOMYTNVDKDLVGMWPAQGSRLTPTCTCGSSDLY 120

Db 59 TCINGVCWTVYHGAGTGTIASPKGPVTOMYTNVDKDLVGMWPAQGSRLTPTCTCGSSDLY 118

QY 121 LVTRHADVIPVRRRGDSRGLSPRPISYLGKSSGGPLLCPCAGHAGVIFRAAVCTRGVAK 180

Db 119 LVTRHADVIPVRRRGDSRGLSPRPISYLGKSSGGPLLCPCAGHAGVIFRAAVCTRGVAK 178

QY 181 AVDFIPVESLETMRSP 197

Db 179 AVDFIPVESLETMRSP 195

RESULT 9

AAY24943

ID AAY24943 standard; Protein; 665 AA.

XX AAY24943;

XX 07-SEP-1999 (first entry)

XX HCV NS4A-NS3 complex SEQ ID NO:14.

KW HCV; hepatitis C virus; single chain recombinant complex; linker;

KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;

KW hydrophobic domain; covalent complex; detection; inhibitor.

XX Hepatitis C virus.

OS Synthetic.

XX WO9928482-A2.

XX 10-JUN-1999.

XX 24-NOV-1998; 98WO-US24528.

XX 28-JUL-1998; 98US-0094331.

XX 28-NOV-1997; 97US-0067315.

XX (SCHE) SCHERING CORP.
 XX PA Malcolm BA, Taremi SS, Weber PC, Yao N;
 XX PI MPI; 1999-385385/32.
 XX DR New hepatitis C virus covalent complexes
 XX PT Claim 6; Page 90-92; 21lpp; English.
 XX PS

XX The present invention describes a covalent hepatitis C virus (HCV)
 CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
 CC to the amino terminus of the HCV NS3 protease domain. The present
 CC sequence represents a specifically claimed example of the above
 CC complex. The covalent NS4A-NS3 complexes are useful for structural
 CC determination and determination of mode of binding of HCV inhibitors by
 CC NMR spectroscopy. They can also be used for detecting inhibitors of the
 CC protease activity, the helicase activity and the ATPase activity of NS3.
 CC The covalent NS4A-NS3 complexes are more soluble, stable and active than
 CC the non-covalent protease-peptide complexes previously available.
 XX

XX Sequence 665 AA;

Query Match 88.3%; Score 911.5; DB 20; Length 665;
 Best Local Similarity 86.7%; Pred. No. 1.9e-87;
 Matches 170; Conservative 16; Mismatches 7; Indels 3; Gaps 1;
 QY 5 GSVVIVGRINLSGD---TAYAAQTRGEGCOETSGTRDKNOVEGEVOIVSTATQTFLAT 61
 DB 22 GSVVIVGRILSGSGSITAYSQOTRGLLGCKKTSLTGRDKNOVEGEVOIVSTATQSFAT 81
 QY 62 CINGVCWTVYHGAGTTRTASPKGPVTOMYTNVDKDLVGHQAPQGSRLTPTCTCGSSDLYL 121
 DB 82 CVNGVCWTVYHGAGSKTLAGPKGPITOMYTNVDQDLVGHQAPPGARSLSLTPTCTCGSSDLYL 141
 QY 122 VTRHADVIPVRRRGDSRGSLLSPRISYLGKSGSGPLLCPCAGHAVGIFRAAVCTRGVAKA 181
 DB 142 VTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSGGPLLCPCSGHAVGIFRAAVCTRGVAKA 201
 QY 182 VDFIPVESLETMRSP 197
 DB 202 VDFVPVESMETMRSP 217

RESULT 10
 AAY24947
 ID AAY24947 standard; Protein; 665 AA.
 XX AC AAY24947;
 XX DT 07-SEP-1999 (first entry)
 XX DE HCV NS4A-NS3 complex SEQ ID NO:18.

XX HCV; hepatitis C virus; single chain recombinant complex; linker;
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
 KW hydrophobic domain; covalent complex; detection; inhibitor.

XX Hepatitis C virus.
 OS Synthetic.
 XX WO9928482-A2.
 XX PD 10-JUN-1999.
 XX PF 24-NOV-1998; 98WO-US24528.
 XX PR 28-JUL-1998; 98US-0094331.
 XX PR 28-NOV-1997; 97US-0067315.

PA (SCHE) SCHERING CORP.
 XX Malcolm BA, Taremi SS, Weber PC, Yao N;
 XX PI MPI; 1999-385385/32.
 XX DR New hepatitis C virus covalent complexes
 XX PT Claim 6; Page 100-102; 21lpp; English.
 XX PS

XX The present invention describes a covalent hepatitis C virus (HCV)
 CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
 CC to the amino terminus of the HCV NS3 protease domain. The present
 CC sequence represents a specifically claimed example of the above
 CC complex. The covalent NS4A-NS3 complexes are useful for structural
 CC determination and determination of mode of binding of HCV inhibitors by
 CC NMR spectroscopy. They can also be used for detecting inhibitors of the
 CC protease activity, the helicase activity and the ATPase activity of NS3.
 CC The covalent NS4A-NS3 complexes are more soluble, stable and active than
 CC the non-covalent protease-peptide complexes previously available.
 XX

XX Sequence 665 AA;

Query Match 88.0%; Score 908.5; DB 20; Length 665;
 Best Local Similarity 86.2%; Pred. No. 3.9e-87;
 Matches 169; Conservative 17; Mismatches 7; Indels 3; Gaps 1;
 QY 5 GSVVIVGRINLSGD---TAYAAQTRGEGCOETSGTRDKNOVEGEVOIVSTATQTFLAT 61
 DB 22 GSVVIVGRILSGSGSITAYSQOTRGLLGCKKTSLTGRDKNOVEGEVOIVSTATQSFAT 81
 QY 62 CINGVCWTVYHGAGTTRTASPKGPVTOMYTNVDKDLVGHQAPQGSRLTPTCTCGSSDLYL 121
 DB 82 CVNGVCWTVYHGAGSKTLAGPKGPITOMYTNVDQDLVGHQAPPGARSLSLTPTCTCGSSDLYL 141
 QY 122 VTRHADVIPVRRRGDSRGSLLSPRISYLGKSGSGPLLCPCAGHAVGIFRAAVCTRGVAKA 181
 DB 142 VTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSGGPLLCPCSGHAVGIFRAAVCTRGVAKA 201
 QY 182 VDFIPVESLETMRSP 197
 DB 202 VDFVPVESMETMRSP 217

RESULT 11
 AAY24941
 ID AAY24941 standard; Protein; 665 AA.
 XX AC AAY24941;
 XX DT 07-SEP-1999 (first entry)
 XX DE HCV NS4A-NS3 complex SEQ ID NO:12.

XX HCV; hepatitis C virus; single chain recombinant complex; linker;
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
 KW hydrophobic domain; covalent complex; detection; inhibitor.

XX Hepatitis C virus.
 OS Synthetic.
 XX WO9928482-A2.
 XX PD 10-JUN-1999.
 XX PF 24-NOV-1998; 98WO-US24528.
 XX PR 28-JUL-1998; 98US-0094331.
 XX PR 28-NOV-1997; 97US-0067315.

(SCHE) SCHERING CORP.

XX MalcolM BA, Taremi SS, Weber PC, Yao N;
 PI WPI; 1999-385385/32.
 DR New hepatitis C virus covalent complexes
 XX Claim 6; Page 85-87; 21lpp; English.

XX The present invention describes a covalent hepatitis C virus (HCV)
 CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
 CC to the amino terminus of the HCV NS3 protease domain. The present
 CC sequence represents a specifically claimed example of the above
 CC complex. The covalent NS4A-NS3 complexes are useful for structural
 CC determination and determination of mode of binding of HCV inhibitors by
 CC NMR spectroscopy. They can also be used for detecting inhibitors of the
 CC protease activity, the helicase activity and the ATPase activity of NS3.
 CC The covalent NS4A-NS3 complexes are more soluble, stable and active than
 CC the non-covalent protease-peptide complexes previously available.

XX Sequence 665 AA;

Query Match 87.9%; Score 907.5; DB 20; Length 665;
 Best Local Similarity 86.7%; Pred. No. 5e-87;
 Matches 170; Conservative 15; Mismatches 8; Indels 3; Gaps 1;

QY 5 GSVVIVGRINLSGD---TAYAAQOTRGECCOETSOTGRKNOVEGEVOIVSTATOTFLAT 61
 DB 22 GSVVIVGRILLSGSGSIYAYSOQTRGLLGCKITSLTGRKNOVEGEVOIVSTATOSFLAT 81
 QY 62 CINGVCWTYVYHGAGTRTIASPKGPVTOMYTNVDKDLVGMQAPQGSRLTPTCTCGSSDLYL 121
 DB 82 CVNGVCWTYVYHGAGSKTLAGPKGPITOMYTNVDQDLVGMQAPPGARSILPTCTCGSSDLYL 141
 QY 122 VTRHADVIPVRRRGDSRGLSPRPISYKLGSSGGPILCPAGHAVGIFRAAVCTRGVAKA 181
 DB 142 VTRHADVIPVRRRGDSRGLSPRPVSYKLGSSGGPILCPSGHAVGIFRAAVCTRGVAKA 201
 QY 182 VDFIPVESLETTMRSP 197
 DB 202 VDFVPVESMETMRSP 217

RESULT 12
 AAY24942
 ID AAY24942 standard; Protein; 665 AA.

XX AAY24942;

XX 07-SEP-1999 (first entry)
 XX HCV NS4A-NS3 complex SEQ ID NO:13.

XX HCV; hepatitis C virus; single chain recombinant complex; linker;
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
 KW hydrophobic domain; covalent complex; detection; inhibitor.

OS Hepatitis C virus.
 OS Synthetic.

XX WO9928482-A2.

XX 10-JUN-1999.

XX 24-NOV-1998; 98WO-US24528.

XX 28-JUL-1998; 98US-0094331.

XX 28-NOV-1997; 97US-0067315.

XX (SCHE) SCHERING CORP.

XX

PI MalcolM BA, Taremi SS, Weber PC, Yao N;
 DR WPI; 1999-385385/32.
 XX New hepatitis C virus covalent complexes
 XX Claim 6; Page 88-90; 21lpp; English.

XX The present invention describes a covalent hepatitis C virus (HCV)
 CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
 CC to the amino terminus of the HCV NS3 protease domain. The present
 CC sequence represents a specifically claimed example of the above
 CC complex. The covalent NS4A-NS3 complexes are useful for structural
 CC determination and determination of mode of binding of HCV inhibitors by
 CC NMR spectroscopy. They can also be used for detecting inhibitors of the
 CC protease activity, the helicase activity and the ATPase activity of NS3.
 CC The covalent NS4A-NS3 complexes are more soluble, stable and active than
 CC the non-covalent protease-peptide complexes previously available.

XX Sequence 665 AA;

Query Match 87.9%; Score 907.5; DB 20; Length 665;
 Best Local Similarity 86.7%; Pred. No. 5e-87;
 Matches 170; Conservative 15; Mismatches 8; Indels 3; Gaps 1;

QY 5 GSVVIVGRINLSGD---TAYAAQOTRGECCOETSOTGRKNOVEGEVOIVSTATOTFLAT 61
 DB 22 GSVVIVGRILLSGSGSIYAYSOQTRGLLGCKITSLTGRKNOVEGEVOIVSTATOSFLAT 81
 QY 62 CINGVCWTYVYHGAGTRTIASPKGPVTOMYTNVDKDLVGMQAPQGSRLTPTCTCGSSDLYL 121
 DB 82 CVNGVCWTYVYHGAGSKTLAGPKGPITOMYTNVDQDLVGMQAPPGARSILPTCTCGSSDLYL 141
 QY 122 VTRHADVIPVRRRGDSRGLSPRPISYKLGSSGGPILCPAGHAVGIFRAAVCTRGVAKA 181
 DB 142 VTRHADVIPVRRRGDSRGLSPRPVSYKLGSSGGPILCPSGHAVGIFRAAVCTRGVAKA 201
 QY 182 VDFIPVESLETTMRSP 197
 DB 202 VDFVPVESMETMRSP 217

RESULT 13
 AAY17880
 ID AAY17880 standard; Protein; 216 AA.

XX AAY17880;

XX 07-SEP-1999 (first entry)

XX HCV NS4A-NS3 complex SEQ ID NO:4.

XX HCV; hepatitis C virus; single chain recombinant complex; linker;
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
 KW hydrophobic domain; covalent complex; detection; inhibitor.

OS Hepatitis C virus.
 OS Synthetic.

XX WO9928482-A2.

XX 10-JUN-1999.

XX 24-NOV-1998; 98WO-US24528.

XX 28-JUL-1998; 98US-0094331.

XX 28-NOV-1997; 97US-0067315.

XX (SCHE) SCHERING CORP.

XX MalcolM BA, Taremi SS, Weber PC, Yao N;

```
XX DR WPI; 1999-385385/32.
XX PT New hepatitis C virus covalent complexes
XX PS Claim 6; Page 76-77; 21pp; English.
XX CC The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence represents a specifically claimed example of the above
CC complex. The covalent NS4A-NS3 complexes are useful for structural
CC determination and determination of mode of binding of HCV inhibitors by
CC NMR spectroscopy. They can also be used for detecting inhibitors of the
CC protease activity, the helicase activity and the ATPase activity of NS3.
CC The covalent NS4A-NS3 complexes are more soluble, stable and active than
CC the non-covalent protease-peptide complexes previously available.
XX SQ Sequence 216 AA;
Query Match 87.6%; Score 904.5; DB 20; Length 216;
Best Local Similarity 86.7%; Pred. No. 2.2e-87;
Matches 169; Conservative 16; Mismatches 7; Indels 3; Gaps 1;
OY 5 GSVVIVGRINLSGD---TAYAQOTRGEQCOFTSQTGRDNQVEGEVQIYSTATQTFLAT 61
DB 22 GSVVIVGRILSGSGSITAYSQOTRGLGCKKTSLTGRDNQVEGEVQVYSTATQSFAT 81
OY 62 CINGVCWTVYHGACGTRTIASPKGPVTOMYTNVDKLVGWQAPOGSRLTPTCTGSSDLYL 121
DB 82 CVNGVCWTVYHGAGSKTLAGPKGPITQMTYTNVDQDLVGWQAPPQARSLLTPTCTGSSDLYL 141
OY 122 VTRHADVIPVRRRGDSRGSLLSPRISYLGSGSGGLLCPAGHAGVGFRAAVCTRGVAKA 181
DB 142 VTRHADVIPVRRRGDSRGSLLSPRISYLGSGSGGLLCPAGHAGVGFRAAVCTRGVAKA 201
OY 182 VDFIPVESLETTMRS 196
DB 202 VDFVPVESMETTMRSP 216
RESULT 14
AAY24945
ID AAY24945 standard; Protein: 665 AA.
XX AC AAY24945;
XX DT 07-SEP-1999 (first entry)
XX DE HCV NS4A-NS3 complex. SEQ ID NO:16.
XX KW HCV; hepatitis C virus; single chain recombinant complex; linker;
XX NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
XX hydrophobic domain; covalent complex; detection; inhibitor.
XX OS Hepatitis C virus.
XX OS Synthetic.
XX PN W09928482-A2.
XX PD 10-JUN-1999.
XX PF 24-NOV-1998; 98WO-US24528.
XX PR 28-JUL-1998; 98US-0094331.
XX PR 28-NOV-1997; 97US-0067315.
XX PA (SCHE ) SCHERING CORP.
XX PI Malcolm BA, Taremi SS, Weber PC, Yao N;
XX DR WPI; 1999-385385/32.
XX PT New hepatitis C virus covalent complexes
XX PS Claim 6; Page 95-97; 21pp; English.
XX CC The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence represents a specifically claimed example of the above
CC complex. The covalent NS4A-NS3 complexes are useful for structural
CC determination and determination of mode of binding of HCV inhibitors by
CC NMR spectroscopy. They can also be used for detecting inhibitors of the
CC protease activity, the helicase activity and the ATPase activity of NS3.
CC The covalent NS4A-NS3 complexes are more soluble, stable and active than
CC the non-covalent protease-peptide complexes previously available.
XX SQ Sequence 665 AA;
Query Match 87.6%; Score 904.5; DB 20; Length 665;
Best Local Similarity 86.2%; Pred. No. 1e-86;
Matches 169; Conservative 16; Mismatches 8; Indels 3; Gaps 1;
OY 5 GSVVIVGRINLSGD---TAYAQOTRGEQCOFTSQTGRDNQVEGEVQIYSTATQTFLAT 61
DB 22 GSVVIVGRILSGSGSITAYSQOTRGLGCKKTSLTGRDNQVEGEVQVYSTATQSFAT 81
OY 62 CINGVCWTVYHGACGTRTIASPKGPVTOMYTNVDKLVGWQAPOGSRLTPTCTGSSDLYL 121
DB 82 CVNGVCWTVYHGAGSKTLAGPKGPITQMTYTNVDQDLVGWQAPPQARSLLTPTCTGSSDLYL 141
OY 122 VTRHADVIPVRRRGDSRGSLLSPRISYLGSGSGGLLCPAGHAGVGFRAAVCTRGVAKA 181
DB 142 VTRHADVIPVRRRGDSRGSLLSPRISYLGSGSGGLLCPAGHAGVGFRAAVCTRGVAKA 201
OY 182 VDFIPVESLETTMRS 197
DB 202 VDFVPVESMETTMRSP 217
RESULT 15
AAY24946
ID AAY24946 standard; Protein: 665 AA.
XX AC AAY24946;
XX DT 07-SEP-1999 (first entry)
XX DE HCV NS4A-NS3 complex. SEQ ID NO:17.
XX KW HCV; hepatitis C virus; single chain recombinant complex; linker;
XX NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
XX hydrophobic domain; covalent complex; detection; inhibitor.
XX OS Hepatitis C virus.
XX OS Synthetic.
XX PN W09928482-A2.
XX PD 10-JUN-1999.
XX PF 24-NOV-1998; 98WO-US24528.
XX PR 28-JUL-1998; 98US-0094331.
XX PR 28-NOV-1997; 97US-0067315.
XX PA (SCHE ) SCHERING CORP.
XX PI Malcolm BA, Taremi SS, Weber PC, Yao N;
XX DR WPI; 1999-385385/32.
```

```

XX      New hepatitis C virus covalent complexes
XX
XX      Claim 6; Page 97-99; 21lpp; English.
XX
XX      The present invention describes a covalent hepatitis C virus (HCV)
XX      NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
XX      NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
XX      hydrophobic domain of native HCV NS4A peptide is tethered by the linker
XX      to the amino terminus of the HCV NS3 protease domain. The present
XX      sequence represents a specifically claimed example of the above
XX      complex. The covalent NS4A-NS3 complexes are useful for structural
XX      determination and determination of mode of binding of HCV inhibitors by
XX      NMR spectroscopy. They can also be used for detecting inhibitors of the
XX      protease activity, the helicase activity and the ATPase activity of NS3.
XX      The covalent NS4A-NS3 complexes are more soluble, stable and active than
XX      the non-covalent protease-peptide complexes previously available.
XX
SQ      Sequence 665 AA;
      Query Match      87.6%; Score 904.5; DB 20; Length 665;
      Best Local Similarity 86.2%; Pred. No. 1e-86;
      Matches 169; Conservative 16; Mismatches 8; Indels 3; Gaps 1;
QY      5 GSVVIVGRINLSGD---TAYAQOTRGEQCOETQOTGRDKKNQVEGEVQIVSTATQTFLAT 61
Db      ||||||| ||| ||| ||||| || ||| ||||||| ||||| ||||| |||||
      22 GSVVIVGRILSGSGSITAYSQOTRGLLGC|K|TSLTGRDKKNQVEGEVQVVSATQSFLAT 81
QY      62 CINGVCWTYYHGAGTRTTIASPKGPVTQMYTNVDRKDLVGVQAPQGSRLTPCTCGSSDLYL 121
Db      |:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      82 CVNGVCWTYYHGAGSKTLAGPKGPITQMYTNVDQDLVGVQAPPGARS|LPCTCGSSDLYL 141
QY      122 VTRHADVIPVRRRGDSRGSLLSPRISYLGSSGGPLLC|PAGHAVGIFRAAVCTRGVAKA 181
Db      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
      142 VTRHADVIPVRRRGDSRGSLLSPRISYLGSSGGPLLC|PAGHAVGIFRAAVCTRGVAKA 201
QY      182 VDFIPVESLETMRSP 197
Db      |||:||||| |||||
      202 VDFVPVESMETMRSP 217
```

Search completed: August 30, 2003, 19:12:23
Job time : 44.6227 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 18:01:52 ; Search time 9.75674 Seconds
(without alignments)
949.524 Million cell updates/sec

Title: US-09-965-594-16

Perfect score: 1032

Sequence: 1 MKKGSWIVGRINLSGDTA.....VAKAVDFIPVESLETMRSP 197

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	884.5	85.7	3011	1	POLG_HCV1
2	878.5	85.1	3011	1	POLG_HCVH
3	867.5	84.1	3010	1	POLG_HCVTW
4	857.5	83.1	3010	1	POLG_HCVJT
5	853.5	82.7	3010	1	POLG_HCVBK
6	853.5	82.7	3010	1	POLG_HCVJA
7	677	65.6	3033	1	POLG_HCVJ6
8	675	65.4	3033	1	POLG_HCVJ8
9	86	8.3	321	1	HHOA_ARATH
10	84.5	8.2	485	1	Y136_TREPA
11	81	7.8	452	1	AAMP_HUMAN
12	78.5	7.6	437	1	DEGL_ARATH
13	78	7.6	1705	1	PTPO_MOUSE
14	76.5	7.4	264	1	CTRL_HUMAN
15	76.5	7.4	323	1	VPRT_SMRVH
16	76.5	7.4	333	1	MOSA_RHIME
17	76	7.4	911	1	TB11_NEIMB
18	75.5	7.3	209	1	PAAD_PSEAE
19	75.5	7.3	2663	1	CENE_HUMAN
20	75	7.3	388	1	ODPT_HUMAN
21	75	7.3	455	1	TMS5_MOUSE
22	75	7.3	594	1	NIR_SPIOI
23	74.5	7.2	706	1	TRFE_HORSE
24	74.5	7.2	764	1	ICCR_DROME
25	74	7.2	844	1	CN4A_RAT
26	73.5	7.1	263	1	GRAK_MOUSE
27	73	7.1	259	1	IBPL_HUMAN
28	72.5	7.0	452	1	MTD_ECOLI
29	72.5	7.0	2768	1	THYG_RAT
30	72	7.0	349	1	TRPD_PSEPU
31	72	7.0	387	1	GALL_STRCO
32	72	7.0	1165	1	POL_GALV
33	72	7.0	1210	1	EGFR_MOUSE

34	71.5	6.9	248	1	GRAD_MOUSE
35	71.5	6.9	248	1	TRY1_CHICK
36	71.5	6.9	257	1	GRAM_HUMAN
37	71.5	6.9	355	1	CMG2_SCHPO
38	71.5	6.9	415	1	2P3_RABIT
39	71	6.9	326	1	PANE_RHILO
40	71	6.9	730	1	HELS_METMA
41	71	6.9	915	1	TBP1_MEIGO
42	70.5	6.8	478	1	MM03_RABIT
43	70.5	6.8	642	1	ENV_FIVGL
44	70.5	6.8	1399	1	RPOC_PSEAE
45	70	6.8	397	1	GALL_STRLI

ALIGNMENTS

RESULT 1
POLG_HCV1
ID POLG_HCV1 STANDARD; PRT: 3011 AA.
AC P26664;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
DE Hepatitis C virus (isolate 1) (HCV).
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11104;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91172826; PubMed=1848704;
RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,
RA Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J.,
RA Bradley D.W., Kuo G., Houghton M.;
RA *Genetic organization and diversity of the hepatitis C virus.*;
RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).
CC -I- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -I- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -I- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
CC {RNA}(N).
CC -I- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -I- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M62321; AAA45676.1; -
CC PIR: A39166; GNMVC3.
CC PDB: 1AIV; 16-FEB-99.
CC PDB: 1HEI; 25-NOV-98.
CC MEROPS: S29.001; -
CC MEROPS: U39.001; -
CC InterPro: IPR001410; DEAD.
CC InterPro: IPR002522; HCV_capsid.
CC -----

DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002186; HCV_RdRP.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; Helicase_C; 1.
 DR Pfam: PF00398; Viral_RdRP; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR PolyProtein: Glycoprotein; Transferase: RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 1007 1615 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4 (POTENTIAL).
 FT CHAIN 2014 3011 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT TRANSEM 347 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NF_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 476 476 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;

Query Match 85.78; Score 884.5; DB 1; Length 3011;
 Best Local Similarity 84.38; Pred. No. 4.3e-76;
 Matches 172; Conservative 9; Mismatches 14; Indels 9; Gaps 1;
 3 KKGSWIVGRIN-----LSGDTAYAAQTRGEGCGQTSOTGRKKNQVEGEVIVST 53

Db 1005 RRGREILGPADGWSKGNRLLAPITAYAAQTRGILGCIITSLTGRDNQVEGEVIVST 1064
 QY 54 ATOTFLATCINGCVTVYHAGTRTITASPKGPVTOMYTNVDKDLVGMQAPQGRSLTPTCT 113
 Db 1065 AAOTFLATCINGCVTVYHAGTRTITASPKGPVTOMYTNVDQDLVGMQAPQGRSLTPTCT 1124
 QY 114 CGSSDLYLVRHADVIPVRRGDSRGLSPISYLVKSGSGPILCPAGHAGVIFRAAV 173
 Db 1125 CGSSDLYLVRHADVIPVRRGDSRGLSPISYLVKSGSGPILCPAGHAGVIFRAAV 1184
 QY 174 CTRGAKAVDFIPVESLETMRSP 197
 Db 1185 CTRGAKAVDFIPVENLETMRSP 1208
 RESULT 2
 POLG_HCVH STANDARD; PRT: 3011 AA.
 ID POLG_HCVH AC P27958;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate H) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=111108;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9205256; PubMed=1658800;
 RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
 RA Prince A.M.;
 RT "Genomic structure of the human prototype strain H of hepatitis C
 RT virus: comparison with American and Japanese isolates.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
 RX MEDLINE=97331322; PubMed=9187654;
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
 RA "Structure of the hepatitis C virus RNA helicase domain.";
 RL Nat. Struct. Biol. 4:463-467(1997).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
 RX MEDLINE=98154321; PubMed=9493270;
 RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
 RA Murcko M.A., Lin C., Caron P.R.;
 RT "Hepatitis C virus NS3 RNA helicase domain with a bound
 RT oligonucleotide: the crystal structure provides insights into the mode
 RT of unwinding.";
 RL Structure 6:89-100(1998).
 CC -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
 CC -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
 CC -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
 CC ACTIVATION OF NS3.
 CC -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
 CC -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the p6
 CC position, Cys or Thr in p1 and Ser or Ala in p1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC {RNA}(N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.

Matches 170; Conservative 10; Mismatches 15; Indels 9; Gaps 1;

QY 3 KGSVVIVGRIN-----LSGDTAYAQOTRGECCQETSGTGRKNQVEGEVQIVST 53
DB 1005 RRGQELLGPADGMYKSGWRLLAPITAYAQOTRGLLGCITISLTGRKNQVEGEVQIVST 1064
QY 54 ATOTELATCINGCVTVYHGAGTRTASPKGPVTQMYTNVDKLVGMQAPGSGSRLTPCT 113
DB 1065 ATOTELATCINGCVTVYHGAGTRTASPKGPVTQMYTNVDKLVGMQAPGSGSRLTPCT 1124
QY 114 CGSSDLYLVTRADVIPVRRGDSRGSLLSPRISYLGSSGGPILCPAGHAGVIFRAAV 173
DB 1125 CGSSDLYLVTRADVIPVRRGDSRGSLLSPRISYLGSSGGPILCPAGHAGVIFRAAV 1184
QY 174 CTRGVAKAVDFIPVESLETTRSP 197
DB 1185 CTRGVAKAVDFIPVENLETTRSP 1208

RESULT 3
POLG_HCVTW STANDARD; PRT; 3010 AA.

AC P29846;
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
(GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
(EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
(EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate Taiwan) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31645;
RN [1]

SEQUENCE FROM N.A.
RA MEDLINE=92230206; PubMed=1314449;
RX Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;
RT "The Taiwanese hepatitis C virus genome: sequence determination and
mapping the 5' terminus of viral genomic and antigenomic RNA.";
RL Virology 188:102-113(1992).

CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
precursor polyprotein, commonly with Asp or Glu in the P6
position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
[RNA](N).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).

DB EMBL: M84754; ; NOT_ANNOTATED_CDS.
DR PIR: A40244; GNMVTW.
DR PDB: 1N63; 25-FEB-03.
DR PDB: 1NS3; 08-APR-98.
DR MEROPS: S29.001; .
DR MEROPS: U39.001; .
DR InterPro: IPR001410; DEAD.

DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_env.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRp.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RdRp; 1.
DR Pfam: PF00998; Viral_RdRp; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
DR PolyProtein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
CELLULAR AMINOPEPTIDASE.
FT CHAIN 1 115 CORE PROTEIN (POTENTIAL).
FT CHAIN 116 191 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT CHAIN 2014 3010 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT TRANSMEM 347 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDFE215 CRC64;

Query Match 84.1%; Score 867.5; DB 1; Length 3010;
Best Local Similarity 79.9%; Pred. No. 1.8e-74;
Matches 163; Conservative 18; Mismatches 14; Indels 9; Gaps 1;

```
QY 3 KGSVVIVGRIN-----LSGDTAYAAQTRGEGCQETSQTGRDNQVEGEVQIVST 53
DB 1005 RRGREITLLGADSLGGRWLLAPITAYAAQTRGLFGCIITSLTGRDNQVEGEVQIVST 1064
QY 54 ATQTFLATCINGVCTVYHGAGTRIASPKGPVTOYINVDKDLVQWAPQGSRLTPTCT 113
DB 1065 ATQSFATCINGVCTVYHGAGSKTLAGPKGPITQYINVDODLVGMHAPQGANSLTPTCT 1124
QY 114 CGSSDLXLVTRHADVIPVRRRDSRGLSPRPISYLVKSSGGPLLCFSGHVGIFRAAV 173
DB 1125 CGSSDLXLVTRHADVIPVRRRDSRGLSPRPISYLVKSSGGPLLCFSGHVGIFRAAV 1184
QY 174 CTRGVAKAVDFIPVESLETMRSP 197
DB 1185 CTRGVAKAVDFIPVESLETMRSP 1208

RESULT 4
POLG_HCVJT STANDARD; PRT: 3010 AA.
AC Q00269;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-JT) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31642;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92295714; PubMed=1318627;
RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,
RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
RT "Molecular cloning of hepatitis C virus genome from a single Japanese
RT carrier: sequence variation within the same individual and among
RT infected individuals.";
RL Virus Res. 23:39-53(1992).
CC -I- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -I- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -I- CATALYTIC ACTIVITY: N nucleoside triphosphate ~ N diphosphate +
CC (RNA)(N).
CC -I- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND PROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -I- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D11168; BAA01943.1; -.
CC PIR: A45573; A45573.
CC PDB: 1A1Q; 25-MAR-98.
CC PDB: 1JXP; 14-JAN-98.
CC MEROPS: S29.001; -.
CC MEROPS: U39.001; -.
CC InterPro: IPR001410; DEAD.
QY 3 KGSVVIVGRIN-----LSGDTAYAAQTRGEGCQETSQTGRDNQVEGEVQIVST 53
```

```
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRP.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; Helicase_C; 1.
DR Pfam: PF00998; Viral_RdRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDc; 1.
DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN 1007 1616 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT CHAIN 2014 3010 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT TRANSMEM 347 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT ACT_SITE 1083 1083 POTENTIAL.
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 3010 AA; 326573 MW; 94A1C77435D642BB CRC64;
```

```
Query Match 83.1%; Score 857.5; DB 1; Length 3010;
Best Local Similarity 78.4%; Pred. No. 1.6e-73;
Matches 160; Conservative 20; Mismatches 15; Indels 9; Gaps 1;
QY 3 KGSVVIVGRIN-----LSGDTAYAAQTRGEGCQETSQTGRDNQVEGEVQIVST 53
```

```

Db 1005 RRGREILLGPADSIQGGWRLAPITAYAAQTRGLLCGCIIVTSLTRDKNQVEGEQVWST 1064
QY 54 ATQTFELATCINGCVWTVYHAGAGRTIASPGVPTQMTYNTDKLVGHQAPQGSRLTPCT 113
Db 1065 ATQSFELATCINGCVWTVYHAGAGRTIASPGVPTQMTYNTDKLVGHQAPQGSRLTPCT 1124
QY 114 CGSDLYLTVTRHADVIVPVRGGDSRGSLSPRTISYLGKSGGGLPLCPAGHAGVIFRAAV 173
Db 1125 CGSDLYLTVTRHADVIVPVRGGDSRGSLSPRTISYLGKSGGGLPLCPAGHAGVIFRAAV 1184
QY 174 CTRGVAKAVDIPVSELTTHRSP 197
Db 1185 CTRGVAKAVDIPVSELTTHRSP 1208

RESULT 5
POLG_HCVBK
ID POLG_HCVBK STANDARD; PRT: 3010 AA.
AC P26663;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein (Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66); (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (Isolate BK) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111105;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-911140698; PubMed-1847440;
RA Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
RA Onishi E., Andoh T., Yoshida I., Okayama H.;
RT "Structure and organization of the hepatitis C virus genome isolated
RT from human carriers."
RL J. Virol. 65:1105-1113(1991).
RN [2]
RP SEQUENCE OF 1487-1500.
RX MEDLINE-96235224; PubMed-8647104;
RA Borowski P., Heiland M., Oehlmann K., Becker B., Kornetky L.;
RT "Non-structural protein 3 of hepatitis C virus inhibits
RT phosphorylation mediated by cAMP-dependent protein kinase."
RL Eur. J. Biochem. 237:611-618(1996).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
RX MEDLINE-97015088; PubMed-8861916;
RA Love R.A., Parge H.E., Wickersham J.A., Hostomsky Z., Habuka N.,
RA Moomaw E.W., Adachi T., Hostomsky Z.;
RT "The crystal structure of hepatitis C virus NS3 proteinase reveals a
RT trypsin-like fold and a structural zinc binding site."
RL Cell 87:331-342(1996).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
RX MEDLINE-98227846; PubMed-9568891;
RA Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
RA Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;
RT "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C
RT virus: a 2.2-A resolution structure in a hexagonal crystal form."
RL Protein sci. 7:837-847(1998).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
CC {RNA}(N).

```

```

CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M58335; AAA72945.1; -.
CC FIR: A36465; GNWVTC.
CC PDB: 1A1Q; 25-MAR-98.
CC PDB: 1JXP; 14-JAN-98.
CC PDB: 1NS3; 08-APR-98.
CC PDB: 1C2P; 15-NOV-00.
CC PDB: 1CSJ; 08-NOV-99.
CC PDB: 1GX5; 09-APR-02.
CC PDB: 1GX6; 10-APR-02.
CC PDB: 1QVY; 26-JUN-00.
CC PDB: 80HM; 20-APR-99.
CC MEROPS: S29.001; -.
CC InterPro: IPR001410; DEAD.
CC InterPro: IPR002522; HCV capsid.
CC InterPro: IPR002521; HCV core.
CC InterPro: IPR002519; HCV env.
CC InterPro: IPR002531; HCV_NSI.
CC InterPro: IPR002518; HCV_NS2.
CC InterPro: IPR004109; HCV_NS3.
CC InterPro: IPR000745; HCV_NS4a.
CC InterPro: IPR001490; HCV_NS4b.
CC InterPro: IPR002868; HCV_NS5a.
CC InterPro: IPR002166; HCV_RdRP.
CC InterPro: IPR007095; RNA_pol_DS_PS.
CC InterPro: IPR007094; RNA_pol_PSVir.
CC Pfam: PF01543; HCV capsid; 1.
CC Pfam: PF01542; HCV core; 1.
CC Pfam: PF01539; HCV env; 1.
CC Pfam: PF01560; HCV_NSI; 1.
CC Pfam: PF01538; HCV_NS2; 1.
CC Pfam: PF02907; HCV_NS3; 1.
CC Pfam: PF01006; HCV_NS4a; 1.
CC Pfam: PF01001; HCV_NS4b; 1.
CC Pfam: PF01506; HCV_NS5a; 1.
CC Pfam: PF00998; Viral_RdRP; 1.
CC ProDom: PD186062; HCV_NSI; 1.
CC SMART: SM00487; DEXDC; 1.
CC Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
CC Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
CC Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
CC 3D-structure.
CC INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
CC CELLULAR AMINOPEPTIDASE.
FT CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT TRANSMEM 347 369 POTENTIAL.
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM.
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM.
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM.
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH BOX.

```

FT CARBOHYD 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1031 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX 1039 1047
FT STRAND 1050 1050
FT STRAND 1059 1063
FT STRAND 1068 1074
FT TURN 1075 1076
FT STRAND 1077 1081
FT HELIX 1082 1085
FT TURN 1086 1087
FT STRAND 1090 1092
FT TURN 1093 1094
FT STRAND 1095 1097
FT STRAND 1101 1103
FT TURN 1104 1107
FT STRAND 1108 1112
FT STRAND 1120 1120
FT STRAND 1122 1122
FT STRAND 1129 1133
FT TURN 1135 1136
FT STRAND 1139 1144
FT STRAND 1149 1157
FT HELIX 1158 1161
FT TURN 1162 1163
FT TURN 1165 1166
FT STRAND 1168 1171
FT TURN 1172 1174
FT STRAND 1175 1186
FT TURN 1187 1188
FT STRAND 1189 1197
FT HELIX 1198 1202
FT TURN 1203 1204
FT STRAND 1204 1204
SQ SEQUENCE 3010 AA; F8422D5ECCDFD9C CRC64;

Query Match 82.7%; Score 853.5; DB 1; Length 3010;
Best Local Similarity 77.9%; Pred No. 3.9e-73;
Matches 159; Conservative 21; Mismatches 15; Indels 9; Gaps 1;
QY 3 KKGWIVGRIN-----LSGDYAAQTRGEGCQTSOTGRDNQVEGEVQIVST 53
DB 1005 RRGKEILLGPADSLRGLRLAPITAYSOOTRGLLGLIITSLTGRDNQVEGEVQVST 1064
QY 54 ATQFLATCINGCWTVYHGAGRTIASPQVPTQMTNWDKLVGHQAPQGSRLTPCT 113
DB 1065 ATQFLATCINGCWTVYHGAGRTIASPQVPTQMTNWDKLVGHQAPQGSRLTPCT 1124
QY 114 CGSSDLXVTRHADVIPVRRGDSRGLSPRPTSYLKSGSGGLPCPAGHANGVIFRAAV 173
DB 1125 CGSSDLXVTRHADVIPVRRGDSRGLSPRPTSYLKSGSGGLPCPAGHANGVIFRAAV 1184
QY 174 CTRGVAKAVDFIPVESLETTMRSP 197
DB 1185 CTRGVAKAVDFIPVESLETTMRSP 1208

RESULT 6

POLG_HCVJA STANDARD; PRT; 3010 AA.
ID POLG_HCVJA AC P26662;
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (HCV).
OS Hepatitis C virus (isolate Japanese) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91085550; PubMed=2175903;
RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
RA Sugimura T., Shimotohno K.;
RT "Molecular cloning of the human hepatitis C virus genome from
RT Japanese patients with non-A, non-B hepatitis.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
RN [2]
RP DISCUSSION OF SEQUENCE.
RX MEDLINE=91192160; PubMed=1849488;
RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraio K.,
RA Ohkoshi S., Shimotohno K.;
RT "Molecular structure of the Japanese hepatitis C viral genome.";
RL FEBS Lett. 280:325-328(1991).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
CC (RNA)(N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D90208; BAAL4233.1; -
CC PIR; A39253; GNMVCJ.
CC HSSP; P26663; LJXP.
CC MEROPS; S29.001; -
CC MEROPS; 039.001; -
CC InterPro: IPR001410; DEAD.
CC InterPro: IPR002522; HCV_capsid.
CC InterPro: IPR002521; HCV_core.
CC InterPro: IPR002519; HCV env.
CC InterPro: IPR002531; HCV NS1.
CC InterPro: IPR002518; HCV NS2.
CC InterPro: IPR004109; HCV_NS3.
CC InterPro: IPR000745; HCV_NS4a.
CC InterPro: IPR001490; HCV_NS4b.
CC InterPro: IPR002868; HCV NS5a.
CC InterPro: IPR002166; HCV_RdRP.

DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.

DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; Helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRp; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; dEXDC; 1.

KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural
 REMOVED FROM CAPSID PROTEIN C BY THE
 INIT_MET 1

FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1089
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 532 532
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2788 2788
 FT SEQUENCE 3010 AA; 327017 MW; AA993794F46DB185 CRC64;

Query Match 82.7%; Score 853.5; DB 1; Length 3010;
 Best Local Similarity 77.0%; Pred. No. 3.9e-73;
 Matches 157; Conservative 23; Mismatches 15; Indels 9; Gaps 1;
 QY 3 KKGSVVIVGRINLSGD-----TAYAAQPTFRGEGCQETSTGRKNQVEGEVQIVST 53
 Db 1005 RRGKEILLGPAUDFCEGWRLLAPITAYSQPTGLGLCIITSLTGRKNQVDGEVQIVLST 1064
 QY 54 ATQTFATCINGCVTVYHGATRTIATSPKGPVQTYNVNDKDLVGMQAPGSRSLTPTCT 113
 Db 1065 ATQSFATCNGVCVTVYHGAGSKTLAGPKGPTQTYNVDDLVGMVAPPGARSMTPTCT 1124
 QY 114 CGSSDLYLTRADVTPVRRRGRSGSLSPRPISYLKSSGGPLLCPCAGHAGVIFRAAV 173
 Db 1125 CGSSDLYLTRADVTPVRRRGRSGSLSPRPISYLKSSGGPLLCPCGHVGVIFRAAV 1184
 QY 174 CTRGVAKAVDFIPVESLETTMRSP 197

Db 1185 CTRGVAKAVDFIPVESMETMRSP 1208

RESULT 7

POLG_HCVJ6
 ID POLG_HCVJ6 STANDARD; PRT: 3033 AA.
 AC P26660;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-J6) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11113;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9204440; PubMed=1638196;
 RA Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H.,
 RA Machida A., Miyakawa Y., Mayumi M.;
 RT "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
 RT from a human carrier: comparison with reported isolates for conserved
 RT and divergent regions".
 RL J. Gen. Virol. 72:2697-2704(1991).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 CC [RNA](N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPID-PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D00944; BAA00792.1; -.
 DR PIR: JQ1303; JQ1303.
 DR MEROPS: S29.001; 1HEI.
 DR MEROPS: S29.001; -.
 DR MEROPS: U39.001; -.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRp.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.

DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS4b; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRp; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Coat protein; Hydrolase; Serine protease.
 FT INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 136 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CARBOHYD 627 627
 FT CARBOHYD 649 649
 FT CARBOHYD 1091 1091
 FT CARBOHYD 2038 2038
 FT CARBOHYD 2811 2811
 SQ SEQUENCE 3033 AA: 329165 MW: F957F5C1A273BE9E CRC64;
 Query Match 65.6%; Score 677; DB 1; Length 3033;
 Best Local Similarity 68.7%; Pred. No. 3e-56;
 Matches 123; Conservative 26; Mismatches 30; Indels 0; Gaps 0;
 QY 19 TAYAQTRGEECCQTSQTRDKNQVEGVQIVSTATQTFATCNGVWTVYHGAGRT 78
 DB 1034 TAYAQTRGELGTIVVSMGTQKTEQAGEIQVLSVTQSFGLTTISVGLVTVYHGAGNKT 1093
 QY 79 IASPKGPVTQMTYNDKDLVGNQAPQGSRLPTCTCGSSDLYLVTRHADVTPVRRGDSR 138
 DB 1094 LAGSRGPVTQMTYSSAEGDVLGWSPSPGRTKSLPECTCGAVDLYLVTRNADVIPARRRDKR 1153
 QY 139 GSLLSPRIYILKSGSGGPLLCAGHAGVIFRAAVCTRGKAVADFIPVESLETMTMRSP 197
 DB 1154 GALLSPRLTLKSGSGGPEVLCPRGHAGVGFRAAVCSRGVAKSIDFIPVEILDIVTRSP 1212
 RESULT 8
 ID POLG_HCVJ8
 AC P26661; STANDARD; PRT: 3033 AA.

DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-J8) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11115;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92230232; PubMed-1314459;
 RA Okamoto H., Kurai K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,
 RA Fukuda S., Tsuda F., Mishihiro S.;
 RT *Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes*;
 RL Virology 188:331-341(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position. Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 CC {RNA}(N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D10988; BAA01761.1;
 CC PIR: A40250; GNMVJ8.
 CC HSSP: P27958; 1HEI.
 CC MEROPS: S29.001;
 CC MEROPS: Q39.001;
 CC InterPro: IPR001410; DEAD.
 CC InterPro: IPR002522; HCV_capsid.
 CC InterPro: IPR002521; HCV_core.
 CC InterPro: IPR002519; HCV_env.
 CC InterPro: IPR002531; HCV_NS1.
 CC InterPro: IPR002518; HCV_NS2.
 CC InterPro: IPR004109; HCV_NS3.
 CC InterPro: IPR000745; HCV_NS4a.
 CC InterPro: IPR001490; HCV_NS4b.
 CC InterPro: IPR002868; HCV_NS5a.
 CC InterPro: IPR002166; HCV_RdRp.
 CC InterPro: IPR007095; RNA_pol_DS_PS.
 CC InterPro: IPR007094; RNA_pol_PSVir.
 CC Pfam: PF01543; HCV_capsid; 1.
 CC Pfam: PF01539; HCV_core; 1.
 CC Pfam: PF01538; HCV_env; 1.
 CC Pfam: PF01560; HCV_NS1; 1.
 CC Pfam: PF01538; HCV_NS2; 1.
 CC Pfam: PF02907; HCV_NS3; 1.
 CC Pfam: PF01006; HCV_NS4a; 1.
 CC Pfam: PF01001; HCV_NS4b; 1.
 CC Pfam: PF01506; HCV_NS5a; 1.
 CC Pfam: PF00998; Viral_RdRp; 1.

```

DR ProDom: PD186062; HCV_NSL; 1.
DR SMART; SM00487; DEXDc; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
FT INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 733
FT CHAIN 734 1010
FT CHAIN 1011 1619
FT CHAIN 1620 1866
FT CHAIN 1867 2017
FT CHAIN 2018 3033
FT TRANSMEM 347 369
FT ACT_SITE 1087 1087
FT ACT_SITE 1111 1111
FT ACT_SITE 1169 1169
FT NP_BIND 1234 1241
FT SITE 1320 1323
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 233 233
FT CARBOHYD 299 299
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 477 477
FT CARBOHYD 534 534
FT CARBOHYD 542 542
FT CARBOHYD 558 558
FT CARBOHYD 578 578
FT CARBOHYD 627 627
FT CARBOHYD 649 649
FT CARBOHYD 1091 1091
FT CARBOHYD 2038 2038
FT CARBOHYD 2359 2359
FT CARBOHYD 2811 2811
SQ SEQUENCE 3033 AA; 330177 MW; 1A173E7E3381FD1A CRC64;

Query Match 65.4%; Score 675; DB 1; Length 3033;
Best local Similarity 69.3%; Pred. No. 4.6e-56;
Matches 124; Conservative 24; Mismatches 31; Indels 0; Gaps 0;

QY 19 TAYAQOTRGECCQTSOTGRDKNOVEGVQIVSTATOTFLATCINGVCMTVYHGAGT 78
Db 1034 TAYTOOTRGLLGAIVSVLTGRDKNEAGOVQVLSVVTOTFLGTSIGVLTVYHGAGNT 1093
QY 79 IASPGPVQMTNVDKLVGQAPQGSRLTCTCGSSDLVLTNRADVIPVRRGDSR 138
Db 1094 LAGPGPVQMTNYSAGBLVGPSPGPKSLDPTCGGAVDLVLTNRADVIPVRRKDDR 1153
QY 139 GSLLSPRISYLSKSGGGLPCPAGHAYGIFRAAVCTRGVAKVDFIPVESLTMRSP 197
Db 1154 GALLSPRLSTLKGSGGGLVLCRSHAVGLFRFAAVCARGVAKSIDFIPVSLDVATTP 1212

RESULT 9
HHOA_ARATH STANDARD; PRT: 321 AA.
AC 09SEL7: 049507;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 15-SEP-2003 (Rel. 42, Last annotation update)
DE Protease HhoA, chloroplast precursor (EC 3.4.21.-).
GN HHOA OR AT4G18370 OR F28J12.30.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

```

```

OC eucoids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Lensch M.H.A., Sokolenko A., Herrmann R.G.;
RT "Identification and Characterization of the chloroplast HhoA protease,
a homolog to the bacterial periplasmic protease HhoA.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=cv. Columbia;
RX MEDLINE=20083488; PubMed=10617198;
RA Meyer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terry N.,
Harris B., Ausorge W., Brandt P., Grivell L., Rieger M.,
Weichselgartner M., de Simone V., Obermaier L., Mache R., Mueller M.,
Kreiss M., Delseny M., Puigdomenech P., Watson M., Schmidheini T.,
Reichert B., Potetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
Vos P., Hohelsel J., Zimmermann W., Wedler H., Ridley P.,
Langham S.-A., McCullagh B., Bilham L., Robben J.,
Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
Weitzenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
Moeljan P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
Berneiser S., Hempel S., Feldpausch M., Lanberth S., Van den Daele H.,
De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,
Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,
Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
Borkova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,
Dose S., de Haan M., Maarse A., Schaefer M., Mueller-Auer S.,
Gabel C., Fuchs M., Fartmann B., Grandrath K., Dauner D., Herzl A.,
Neumann S., Argiou A., Vitale D., Liquori R., Piravandi E.,
Massen O., Quigley F., Clabaud G., Muendlein A., Aubourg S.,
Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
Chedor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
Gibbons T., Weber N., Vandenbol M., Bagues M., Terol J., Torres A.,
Perez-Perez A., Furnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Biele C.,
Fishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
Stonking T., Kalicki J., Graves T., Harmon G., Edwards J.,
Latrelle P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshua C.,
Antonova B., Zidanic M., Strong C., Sun H., Lamar B., Jordan C.,
Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S.,
Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
Chen E., Marra M., Martienssen R., McCombie W.R.;
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
thaliana.";
RL Nature 402:769-777(1999).
RN [3]
RP SEQUENCE OF 72-82; 96-110; 150-159; 178-211 AND 306-320.
RA Schubert M., Peterson U., Funk C., Haas B., Schroeder W.P.,
RA Kieselbach T.;
RT "The chloroplast lumen from Arabidopsis thaliana.";
CC -1- SURCELLULAR LOCATION: Chloroplast; within the thylakoid lumen.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
CC -1- CAUTION: Ref.2 sequences differ from that shown due to erroneous
gene model prediction. AT4G18370 and AT4G18375 were originally
fused into a single gene.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its

```


CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; M95627; AAA68889.1; -
DR PIR; I39383; I39383.
DR Genew; HGNC:18; AAMP.
DR MIM; 603488; -
DR GO; GO:008201; F:heparin binding activity; TAS.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00678; WD_REPEATS.1; 1.
DR PROSITE; PS0082; WD_REPEATS.2; 6.
DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
FT DOMAIN 14 18 HEPARIN-BINDING (POTENTIAL).
FT DOMAIN 71 77 POLY-GLU.
FT REPEAT 107 138 WD 1.
FT REPEAT 150 180 WD 2.
FT REPEAT 190 220 WD 3.
FT REPEAT 231 261 WD 4.
FT REPEAT 276 306 WD 5.
FT REPEAT 333 363 WD 6.
FT REPEAT 374 404 WD 7.
FT REPEAT 416 446 WD 8.
SQ SEQUENCE 452 AA; 49015 MW; DA1413D25E236C0 CRC64;

Query Match 7.8%; Score 81; DB 1; Length 452;
Best Local Similarity 25.3%; Pred. No. 3;
Matches 42; Conservative 13; Mismatches 47; Indels 64; Gaps 9;
QY 68 VYVTHGAGTRTIASPKGPVTQMTYVNDKLVGQAPGGSRL-----TPCTCGSSDLV 122
D 197 WMEWH-----PRAPVLGACT-ADGNTWMKVPNGDKCTFGPNCPTACGR----- 240
QY 123 TRHADVTPVRR---GDSRGS-----LLSPRTSYLKGSSG--GPLLCPA----- 162
D 241 -----VLPDGKRAVGVYEDGIRIWDLKQSPHVLKGTCHGGLTCVAANQDGLILT 295
QY 163 -----GHAVGIFR-----AAVCTRGVAKAVDFIPVESL 190
D 296 GSVDCQAKLVSAITGKVVGVFETVATSPQSLGEGEESNSVESL 341

RESULT 12
DEGLARATH
ID DEGLARATH STANDARD; PRT; 437 AA.
AC 022609; O9LK85;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protease Do-like 1, chloroplast precursor (EC 3.4.21.-).
GN DEGP1 OR DEGP OR AT3G27925 OR K16N12.18.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eudicots II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;

RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=98175982; PubMed=9507020;
RA Itzhaki H., Naveh L., Lindahl M., Cook M., Adam Z.;
RT Identification and characterization of DegP, a serine protease
RT associated with the luminal side of the thylakoid membrane.;
RL J. Biol. Chem. 273:7094-7098(1998).
RN [2]
RP SEQUENCE FROM N.A.
RP STRAIN=cv. Columbia;

RX MEDLINE=20363099; PubMed=10907853;
RA Kaneko T., Katoh T., Sato S., Nakamura A., Asamizu E., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the 4,251,695 bp regions covered by 90 P1, TAC
RT and BAC clones.";
RL DNA Res. 7:217-221(2000).
RN [3]
RP SEQUENCE OF 104-118.
RC STRAIN=cv. Columbia;
RA Kieselbach T., Bystedt M., Schroeder W.P.;
RL Submitted (JUL-2000) to the SWISS-PROT data bank.
CC -!- FUNCTION: SERINE PROTEASE THAT IS REQUIRED AT HIGH TEMPERATURE.
CC MAY BE INVOLVED IN THE DEGRADATION OF DAMAGED PROTEINS. IN VIVO,
CC CAN DEGRADE BETA-CASEIN.
CC -!- ENZYME REGULATION: INHIBITED BY PHENYL METHYL SULFONYL FLUORIDE AND
CC O-PHENANTHROLINE.
CC -!- SUBCELLULAR LOCATION: BOUND TO LUMINAL SIDE OF THE THYLAKOID
CC MEMBRANE.
CC -!- INDUCTION: By heat shock.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AF028842; AAC39436.1; -
DR EMBL; AP000371; BAB02539.1; -
DR EMBL; AP001302; BAB02539.1; JOINED.
DR MEROPS; S01.279; -
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS0106; PDZ; 1.
DR PROSITE; PS0106; PDZ; 1.
KW Hydrolyase; Serine protease; Transit peptide; Chloroplast; Thylakoid.
FT TRANSIT 1 ? CHLOROPLAST (POTENTIAL).
FT CHAIN 104 437 PROTEASE DO-LIKE 1.
FT DOMAIN 152 321 SERINE PROTEASE.
FT DOMAIN 324 421 PDZ.
FT ACT_SITE 171 171 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 201 201 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 280 280 CHARGE RELAY SYSTEM (POTENTIAL).
FT CONFLICT 12 23 HSPSSQLSNST -> SSTLFLHSPSSHL (IN REF.
FT CONFLICT 36 36 V -> I (IN REF. 2).
FT CONFLICT 54 54 P -> S (IN REF. 2).
FT CONFLICT 60 60 G -> R (IN REF. 2).
FT CONFLICT 64 64 G -> D (IN REF. 2).
FT CONFLICT 68 69 LL -> HF (IN REF. 2).
FT CONFLICT 355 355 L -> V (IN REF. 2).
FT CONFLICT 381 381 I -> V (IN REF. 2).
FT CONFLICT 416 416 O -> E (IN REF. 2).
SQ SEQUENCE 437 AA; 46213 MW; 1497B1AB3F5FF2M4 CRC64;
Query Match 7.6%; Score 78.5; DB 1; Length 437;
Best Local Similarity 25.6%; Pred. No. 5;
Matches 44; Conservative 17; Mismatches 56; Indels 55; Gaps 7;
QY 70 VYHAGTRTIASPKGPVTQMY-----TNVDKDLVW-----QA 102
D 150 VPQSGSGFVWDKQGHVITNYHVINGASDLRVLADQTTFDKAVGVDFDQDKDAVLRLA 209
QY 103 PQGSRSLTPTCTGSSDLV-----TRHADVTPVRRGDSGSLSPRI 147

Db 210 PK--NKLRLPIVGVADLLVQKQVFAIGNFGLDHTLTITGVISLRLREIS--SAATGRPI 265
 Qy 148 SYL-----KGSSGGLLCPAGHAVGIFRAAVCTRGVAKAVDF-IPVESL 190
 Db 266 QDVLTQDAAINPGSGGLDSSGLTIGINTAIYSPSGASSGVGFSIPVDTV 317

RESULT 13

PTPO_MOUSE STANDARD; PRT; 1705 AA.
 AC P70289;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Embryonic stem cell protein tyrosine phosphatase precursor
 DE (EC 3.1.3.48) (ES cell phosphatase).
 GN ESP.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Embryonic stem cells;
 RX MEDLINE=97109513; PubMed=8951793;
 RA Lee K., Nichols J., Smith A.;
 RT "Identification of a developmentally regulated protein tyrosine
 RT phosphatase in embryonic stem cells that is a marker of
 RT pluripotential epiblast and early mesoderm.";
 RL Mech. Dev. 59:153-164(1996).
 RN [2]
 RP ERRATUM.
 RA Lee K., Nichols J., Smith A.;
 RL Mech. Dev. 61:213-215(1996).
 CC -!- FUNCTION: MAY PLAY A ROLE IN THE MAINTENANCE OF PLURIPOTENCY.
 CC -!- DOWN-REGULATED DURING DIFFERENTIATION.
 CC -!- CATALYTIC ACTIVITY: Protein tyrosine phosphatase + H(2)O = protein
 CC tyrosine + phosphate.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- DEVELOPMENTAL STAGE: DETECTABLE IN THE EPIBLAST OF OOCYTES AND
 CC THROUGHOUT EARLY MOUSE EMBRYO DEVELOPMENT. IN ADULT, EXPRESSION IS
 CC LOCALIZED IN GONADAL GERM CELLS.
 CC -!- SIMILARITY: Contains 2 protein-tyrosine phosphatase domains.
 CC -!- SIMILARITY: Contains 10 fibronectin type III domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U36488; AAC52868.1; -;
 DR HSP; P18052; LYFO.
 DR MGD; MGI:108027; Esp.
 DR InterPro; IPR003961; FN_III.
 DR InterPro; IPR000387; TYR_phosphatase.
 DR InterPro; IPR000242; Tyr_PP.
 DR Pfam; PF00041; fn3; 7.
 DR Pfam; PF00102; Y_phosphatase; 1.
 DR PRINTS; PR00700; PRTYPHPTASE.
 DR SMART; SM00060; FN3; 8.
 DR SMART; SM00194; PTPC; 1.
 DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
 DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
 DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 2.
 KW Hydrolase; Transmembrane; Repeat; Signal; Glycoprotein.
 FT SIGNAL 1 18
 FT CHAIN 19 1705 EMBRYONIC STEM CELL PROTEIN TYROSINE
 FT PHOSPHATASE.
 FT DOMAIN 19 1077 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1078 1100 POTENTIAL.

FT DOMAIN 1101 1705 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 35 110 FIBRONECTIN TYPE-III 1.
 FT DOMAIN 126 199 FIBRONECTIN TYPE-III 2.
 FT DOMAIN 216 288 FIBRONECTIN TYPE-III 3.
 FT DOMAIN 304 373 FIBRONECTIN TYPE-III 4.
 FT DOMAIN 393 454 FIBRONECTIN TYPE-III 5.
 FT DOMAIN 471 543 FIBRONECTIN TYPE-III 6.
 FT DOMAIN 563 634 FIBRONECTIN TYPE-III 7.
 FT DOMAIN 657 722 FIBRONECTIN TYPE-III 8.
 FT DOMAIN 742 813 FIBRONECTIN TYPE-III 9.
 FT DOMAIN 831 905 FIBRONECTIN TYPE-III 10.
 FT DOMAIN 1150 1418 PROTEIN-TYROSINE PHOSPHATASE 1.
 FT DOMAIN 1469 1700 PROTEIN-TYROSINE PHOSPHATASE 2.
 FT ACT_SITE 1350 1350 BY SIMILARITY.
 FT CARBOHYD 74 74 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 89 89 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 117 117 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 174 174 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 259 259 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 299 299 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 345 345 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 431 431 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 551 551 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 570 570 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 620 620 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 649 649 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 663 663 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 737 737 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 851 851 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 882 882 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 970 970 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 982 982 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 1705 AA; 186795 MW; 2783755F15387D5B CRC64;

Query Match 7.68; Score 78; DB 1; Length 1705;
 Best Local Similarity 25.5%; Pred. No. 28;
 Matches 42; Conservative 12; Mismatches 53; Indels 58; Gaps 8;

QY 9 IVGRINLSGDTAVAAQOTRG-EGCOETTSOTGR-----DKNOVEGEVOIVSTATOT 57
 Db 382 VEGSIWLAESAARMEVPGARLWLEGLKATQGRALLYSVDAPGLLENISVSSGATHV 441
 QY 58 FLATCINGVCWYTHGAGTRTIASPKGPVQMT-----DIASSMGDITQSLTGYTSPLOSLEIRNSPSDLTICW 494
 Db 442 TFCGLVPGAHYRV-----DIASSMGDITQSLTGYTSPLOSLEIRNSPSDLTICW 494
 QY 101 -QAQGSRSRLTPCTCGSSDLYLVTRADVIPVRRGDSRGSLLSP 144
 Db 495 APAP-----GOMEGYKVTWHOD-----GSQSP 517

RESULT 14

CTRL_HUMAN STANDARD; PRT; 264 AA.
 AC P40313;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Chymotrypsin-like protease CTRL-1 precursor (EC 3.4.21.-).
 GN CTRL OR CTRL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94093544; PubMed=8268911;
 RA Larsen F., Solheim J., Kristensen T., Kolsto A.B., Prydz H.;
 RT "A tight cluster of five unrelated human genes on chromosome
 RT 16q22.1";
 RL Hum. Mol. Genet. 2:1589-1595(1993).
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.

```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; X71874; CAAS0710.1; -
DR EMBL; X71877; CAAS0711.1; -
DR PIR; I38136; I38136.
DR HSSP; P00763; IDPO.
DR MEROPS; S01.256; -.
DR Genew; HGNC:2524; CTRL.
DR MIM; 118888; -.
DR GO; GO:0005615; C:extracellular space; TAS.
DR GO; GO:0007386; P:digestion; TAS.
DR GO; GO:0006508; P:proteolysis and peptidolysis; TAS.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser.protease_Try.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal.
FT SIGNAL 1 18 POTENTIAL
FT PROPEP 19 33 ACTIVATION PEPTIDE (POTENTIAL).
FT CHAIN 34 264 CHYMOTRYPSIN-LIKE PROTEASE CTRL-1.
FT ACT_SITE 75 75 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 121 121 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 214 214 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOXYD 114 114 N-LINKED (GLNAC. . .) (POTENTIAL).
FT DISULFID 19 141 BY SIMILARITY.
FT DISULFID 60 76 HY SIMILARITY.
FT DISULFID 155 220 BY SIMILARITY.
FT DISULFID 187 201 BY SIMILARITY.
FT DISULFID 210 239 BY SIMILARITY.
SQ SEQUENCE 264 AA; 28002 MW; 3F629F02FA6DDB4 CRC64;

Query Match 7.4%; Score 76.5; DB 1; Length 264;
Best Local Similarity 25.9%; Pred. No. 4.3;
Matches 35; Conservative 18; Mismatches 51; Indels 31; Gaps 5;

QY 44 VEGEVOIVSTATOTFLATCINGVCWTVYHGAGRTTASPKGPVTVQMYTNVYDKDLVGVQAP 103
DB 118 MNDVTLKLLASPAQYTRISPC-----LASSNEALTEGLTCV---TTGWGRL 163

QY 104 QGSRSLTPCTCGSSDLYLVYTRHADVIPVRRRGDSRGLSPRI-----SYLKSSGG 156
DB 164 SGVGNVTPAHLOQVALPLVT-----VNCROYWGSSITDSMICAGGAGASSCQDSGG 216

QY 157 PLLCPAGHA---VGI 168
DB 217 PLVCQKGNVWLLIGI 231

RESULT 15
VPRT_SMRVH STANDARD; PRT; 323 AA.
AC P21407;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protease (EC 3.4.23.-).
GN PRT.
OS Squirrel monkey retrovirus (SMRV-H) (SMRV-HLB).
OC Viruses; Retrovirdae; Retroviridae; Betaretrovirus.
OX NCBI_TaxID=11856;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RA MEDLINE=95073750; PubMed=3201749;
RA Oda T., Ikeda S., Watanabe S., Hatsushika M., Akiyama K.,
RA Mitsunobu F.;
RT "Molecular cloning, complete nucleotide sequence, and gene structure
RT of the provirus genome of a retrovirus produced in a human
RT lymphoblastoid cell line.";
RL Virology 167:468-476(1988).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A2.
CC -1- SIMILARITY: Contains 1 G-patch domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; M23385; AAA66452.1; ALT_INIT.
DR PIR; B31827; PRLJHD.
DR HSSP; P06968; IEDU.
DR MEROPS; A02.0PM; -.
DR InterPro; IPR001995; Aspprotease_rtrv.
DR InterPro; IPR001969; Aspprotease_site.
DR InterPro; IPR001428; DeoxyUTPase.
DR InterPro; IPR000467; G_patch.
DR Pfam; PF00692; dUTPase; 1.
DR Pfam; PF01585; G_patch; 1.
DR Pfam; PF00077; rvp; 1.
DR SMART; SM00443; G_patch; 1.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
DR PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR PROSITE; PS50174; G_PATCH; 1.
KW Hydrolase; Aspartyl. protease.
FT DOMAIN 275 321 G-PATCH.
FT ACT_SITE 193 193 BY SIMILARITY.
SQ SEQUENCE 323 AA; 35126 MW; 5D6CEA38BA932786 CRC64;

Query Match 7.4%; Score 76.5; DB 1; Length 323;
Best Local Similarity 23.3%; Pred. No. 5.5;
Matches 34; Conservative 16; Mismatches 49; Indels 47; Gaps 5;

QY 42 NOVGEVOIVSTATOTFLATCINGVCWTVYHGAGRTTASPKG-----PVTOMYTN 92
DB 111 NDFEGEIHILSTKDL-----VTIPKGTSLAQIVILPQQINSN 150

QY 93 VDKDLVGWQAPQSGSKSLTPCTCGSSDLYLV---TRHADVIPVRRRGDSRGLL---SPR 145
DB 151 PKPYRGASAP-----GSSDVYVWQQISQQRPTIKLKLNGKLFSGILDGTGADAT 199

QY 146 PTSYLKSSGGPILCPAGHAVGIPRA 171
DB 200 VISYTHWRPNWPLTTVATHLRGIGQA 225

Search completed: August 30, 2003, 19:13:46
Job time : 10.7567 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 19:00:22 : Search time 37.5921 seconds
(without alignments)
1352.314 Million cell updates/sec

Title: US-09-965-594-16
Perfect score: 1032
Sequence: 1 MKKGSVVIVGRINLSGDTA.....VAKAVDFIPVESLETTMRSP 197

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_todent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	902.5	87.5	4040	12 Q9IFH8	Q9ifh8 mucosal dis
2	888.5	86.1	3011	12 Q36579	Q36579 hepatitis c
3	884.5	85.7	2436	12 Q81756	Q81756 hepatitis c
4	884.5	85.7	3011	12 Q9IFB5	Q9ifb5 hepatitis c
5	884.5	85.7	3011	12 Q9EL58	Q9el58 hepatitis c
6	883.5	85.6	3011	12 Q3463	Q3463 hepatitis c
7	881.5	85.4	3011	12 Q36608	Q36608 hepatitis c
8	881.5	85.4	3015	12 Q9PW45	Q9pw45 hepatitis c
9	881.5	85.4	3015	12 Q9PW09	Q9pw09 hepatitis c
10	879	85.2	181	12 Q91RR8	Q91rr8 hepatitis c
11	879	85.2	181	12 Q91RT5	Q91rt5 hepatitis c
12	877	85.0	181	12 Q91RR5	Q91rr5 hepatitis c
13	877	85.0	181	12 Q91RR2	Q91rr2 hepatitis c
14	877	85.0	181	12 Q91RR9	Q91rr9 hepatitis c
15	876	84.9	181	12 Q91RR3	Q91rr3 hepatitis c
16	876	84.9	181	12 Q91RR4	Q91rr4 hepatitis c

17	876	84.9	181	12 Q91RS1	Q91rs1 hepatitis c
18	876	84.9	181	12 Q91RQ8	Q91rq8 hepatitis c
19	876	84.9	181	12 Q91RT1	Q91rt1 hepatitis c
20	876	84.9	181	12 Q91RR0	Q91rr0 hepatitis c
21	875.5	84.8	3011	12 Q36609	Q36609 hepatitis c
22	874	84.7	181	12 Q91RR6	Q91rr6 hepatitis c
23	874	84.7	181	12 Q91RS9	Q91rs9 hepatitis c
24	873	84.6	181	12 Q91RS3	Q91rs3 hepatitis c
25	872.5	84.5	3011	12 Q9DIT6	Q9dit6 hepatitis c
26	872	84.5	181	12 Q91RT4	Q91rt4 hepatitis c
27	872	84.5	181	12 Q91RS8	Q91rs8 hepatitis c
28	872	84.5	181	12 Q91RT3	Q91rt3 hepatitis c
29	872	84.5	181	12 Q91RS5	Q91rs5 hepatitis c
30	872	84.5	181	12 Q91RS7	Q91rs7 hepatitis c
31	872	84.5	181	12 Q91RT0	Q91rt0 hepatitis c
32	872	84.5	181	12 Q91RS2	Q91rs2 hepatitis c
33	871	84.4	181	12 Q91RS6	Q91rs6 hepatitis c
34	870.5	84.4	3010	12 Q9QP61	Q9qp61 hepatitis c
35	870	84.3	181	12 Q91RS4	Q91rs4 hepatitis c
36	869.5	84.3	3010	12 Q68533	Q68533 hepatitis c
37	869	84.2	181	12 Q91RR7	Q91rr7 hepatitis c
38	869	84.2	181	12 Q91RT6	Q91rt6 hepatitis c
39	869	84.2	3011	12 Q36610	Q36610 hepatitis c
40	868.5	84.2	361	12 Q70817	Q70817 hepatitis c
41	868	84.1	181	12 Q91RT8	Q91rt8 hepatitis c
42	867.5	84.1	361	12 Q70818	Q70818 hepatitis c
43	867	84.0	181	12 Q91RR9	Q91rr9 hepatitis c
44	866.5	84.0	3010	12 Q9DTE2	Q9dte2 hepatitis c
45	866.5	84.0	3010	12 Q99AU2	Q99au2 hepatitis c

ALIGNMENTS

RESULT 1

ID	Q9IFH8	PRELIMINARY;	PRT; 4040 AA.
AC	Q9IFH8;		
DT	01-OCT-2000 (Tremblrel. 15, Created)		
DT	01-OCT-2000 (Tremblrel. 15, Last sequence update)		
DT	01-MAR-2003 (Tremblrel. 23, Last annotation update)		
DE	Genome polyprotein.		
OS	Mucosal disease virus.		
OC	Viruses; ssRNA positive-strand viruses, no DNA stage: Flaviviridae;		
OC	Pestivirus.		
OX	NCBI_TaxID=11099;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20323484; Pubmed=10864644;		
RA	Lai V.C., Zhong W., Skelton A., Ingravallo P., Vassilev V.,		
RA	Donis R.O., Hong Z., Lau J.Y.;		
RT	"Generation and characterization of a hepatitis C virus NS3 protease-		
RT	dependent bovine viral diarrhea virus.";		
RL	J. Virol. 74:6339-6347(2000).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RA	Lai V.C.H., Hong Z.;		
RL	Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF268278; AAF82566.1; .		
DR	HSSP; P26663; 1JXP.		
DR	MEROPS; S31.001; .		
DR	InterPro; IPR000280; CDvir_endptsep80.		
DR	InterPro; IPR001410; DEAD.		
DR	InterPro; IPR004109; HCV_NS3.		
DR	InterPro; IPR002166; HCV_RdRP.		
DR	InterPro; IPR001650; Helicase_C.		
DR	InterPro; IPR001005; Myb_DNA_binding.		
DR	InterPro; IPR001568; RNase_T2.		
DR	InterPro; IPR007095; RNA_pol_DS_PS.		
DR	InterPro; IPR007094; RNA_pol_PSVir.		
DR	Pfam; PF02907; HCV_NS3; 1.		
DR	Pfam; PF00271; Helicase_C; 1.		
DR	Pfam; PF00998; Viral_RdRP; 1.		

```

DR PRINTS: PR00729; COVENDOPTASE.
DR SMART: SM00487; DEXDC; 1.
DR SMART: SM00490; HELIC; 1.
DR PROSITE: PS00037; MYB_1; 1.
DR PROSITE: PS05007; RDRP_POSITIVE; 1.
DR PROSITE: PS05021; RDRP_VIRAL; 1.
DR PROSITE: PS05031; RNASE_T2.2; 1.
DR ATP-binding; Helicase; Hydrolase; Nonstructural protein; Polyprotein.
KW RNA-directed RNA polymerase; Transferase.
KW SEQUENCE 4040 AA; 453073 MW; ADS87791D055B9DC CRC64;

Query Match      87.5%; Score 902.5; DB 12; Length 4040;
Best Local Similarity 90.8%; Pred. No. 3e-81;
Matches 177; Conservative 5; Mismatches 10; Indels 3; Gaps 1;

QY 5 GSVIVGRINLSGD---TAYAQOTRGEQCOETSGTRDKNOVEGEVOIVSTATQTFLAT 61
DB 10 GSVIVGRIVLSGSGSITACAAQOTRGLLGCKIITSLTGRDKNOVEGEVOIVSTATQTFLAT 69
QY 62 CINGVCVTYYHGAGTRTIIASPKGPVTOMYTNVDKDLVGMQAPQGSRLTPTCTCGSSDLYL 121
DB 70 CINGVCVTYYHGAGTRTIIASPKGPVIOMYTNVDQDLVGMQAPQGSRLTPTCTCGSSDLYL 129
QY 122 VTRHADVIPVRRGDSRGSLLSPRISYLVKSGSGPGLLCAGHAGVIFRAAVCTRGVAKA 181
DB 130 VTRHANVIPVRRGDSRGSLLSPRISYLVKSGSGPGLLCAGHAGVIFRAAVCTRGVAKA 189
QY 182 VDFIPVESLETTMRS 196
DB 190 VDFIPVENLETTTRS 204

RESULT 2
O36579 ID O36579 PRELIMINARY; PRT: 3011 AA.
AC O36579;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H77;
RX MEDLINE=97373636; PubMed=9228008;
RA Kolykhalov A.A., Agapov E.V., Blight K.J., Mihalik K., Feinstone S.M.,
RA Rice C.M.;
RT "Transmission of hepatitis C by intrahepatic inoculation with
RT transcribed RNA.";
RL Science 277:570-574(1997).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: AF009606; AAB66324.1; -.
DR HSSP: P27958; 1HE7.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NSI.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR000745; HCV_NS4b.
DR InterPro: IPR001490; HCV_NS5a.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRP.
DR InterPro: IPR002166; HCV_RdRP.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.

```

```

DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NSI; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; Helicase_C; 1.
DR Pfam: PF00998; Viral_RdRP; 1.
DR ProDom: PDI86062; HCV_NSI; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS05007; RDRP_POSITIVE; 1.
DR PROSITE: PS05021; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
KW SEQUENCE 3011 AA; 327182 MW; E2E0E809C63C1B9 CRC64;

Query Match      86.1%; Score 888.5; DB 12; Length 3011;
Best Local Similarity 84.3%; Pred. No. 5.4e-80;
Matches 172; Conservative 10; Mismatches 13; Indels 9; Gaps 1;

QY 3 KKGSVVIYGRIN-----LSGDTAYAQOTRGEQCOETSGTRDKNOVEGEVOIVST 53
DB 1005 RQGQILGPDAGMVKWNRLLAPITAYAQOTRGLLGCIITSLTGRDKNOVEGEVOIVST 1064
QY 54 ATQTFLATCINGVCVTYYHGAGTRTIIASPKGPVTOMYTNVDKDLVGMQAPQGSRLTPTCT 113
DB 1065 ATQTFLATCINGVCVTYYHGAGTRTIIASPKGPVTOMYTNVDQDLVGMQAPQGSRLTPTCT 1124
QY 114 CGSSDLYLVTRHADVIPVRRGDSRGSLLSPRISYLVKSGSGPGLLCAGHAGVIFRAAV 173
DB 1125 CGSSDLYLVTRHADVIPVRRGDSRGSLLSPRISYLVKSGSGPGLLCAGHAGVIFRAAV 1184
QY 174 CTRGVAKAVDFIPVESLETTMRSP 197
DB 1185 CTRGVAKAVDFIPVENLETTMRSP 1208

RESULT 3
Q81756 ID Q81756 PRELIMINARY; PRT: 2436 AA.
AC Q81756;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Genome polyprotein (fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]
RP SEQUENCE FROM N.A.
RA Choo Q.-L., Richman K., Han J.;
RT "The nucleotide sequence of the Hepatitis C viral genome.";
RL Submitted (MAY-1990) to the EMBL/GenBank/DBJ databases.
DR EMBL: M32084; AAA45677.1; -.
DR HSSP: P27958; 1A1V.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002531; HCV_NSI.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRP.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01560; HCV_NSI; 1.
DR Pfam: PF01538; HCV_NS2; 1.

```

```

DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RdRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS05057; RDRP_POSITIVE; 1.
DR PROSITE: PS05052; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
FT NON_TER 1
FT NON_TER 2436 2436
SQ SEQUENCE 2436 AA; 264734 MW; DTB9872900BE3125 CRC64;

Query Match 85.7%; Score 884.5; DB 12; Length 2436;
Best Local Similarity 84.3%; Pred. No. 1e-79;
Matches 172; Conservative 9; Mismatches 14; Indels 9; Gaps 1;

QY 3 KGSVIVGRIN-----LSGDTAYAAQOTRGECCQETSOTGRDKNOVEGEVQIVST 53
Db 555 RRGREILLGPADGMVSKGNRLAPITAYAAQOTRGLGCIITSLTGRDKNOVEGEVQIVST 614
QY 54 ATQTFATCINGVCWTVYHGAGTRTASPKGPVQMTYNDKDLVGQAPQGSRLTPCT 113
Db 615 AAQTFATCINGVCWTVYHGAGTRTASPKGPVQMTYNDKDLVGQAPQGSRLTPCT 674
QY 114 CGSSDLVLTTRHADVIPVRRGDSRGLSPRISYLYKSGSGGPLLCPAGHAGVIFRAAV 173
Db 675 CGSSDLVLTTRHADVIPVRRGDSRGLSPRISYLYKSGSGGPLLCPAGHAGVIFRAAV 734
QY 174 CTRGVAKAVDFIPVESLETTMRSP 197
Db 735 CTRGVAKAVDFIPVENLETTMRSP 758

RESULT 4
Q9IF55 ID Q9IF55 PRELIMINARY; PRT: 3011 AA.
AC Q9IF55;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21262212; PubMed=11369872;
RA Lanford R.E., Lee H., Chavez D., Guerra B., Brasky K.M.;
RT "Infectious cDNA clone of the hepatitis C virus genotype 1 prototype
sequence."
RL J. Gen. Virol. 82:1291-1297(2001).
CC -|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: AF271632; AAF81759.1; -.
DR HSSP: P27958; 1A1V.
DR InterPro: IPR000345; CytC_heme_bind.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.

```

```

DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRP.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RdRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS00190; CYTOCHROME_C; 1.
DR PROSITE: PS05057; RDRP_POSITIVE; 1.
DR PROSITE: PS05052; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
KW SEQUENCE 3011 AA; 327124 MW; 2489CE74AC864E58 CRC64;

Query Match 85.7%; Score 884.5; DB 12; Length 3011;
Best Local Similarity 84.3%; Pred. No. 1.4e-79;
Matches 172; Conservative 9; Mismatches 14; Indels 9; Gaps 1;

QY 3 KGSVIVGRIN-----LSGDTAYAAQOTRGECCQETSOTGRDKNOVEGEVQIVST 53
Db 1005 RRGREILLGPADGMVSKGNRLAPITAYAAQOTRGLGCIITSLTGRDKNOVEGEVQIVST 1064
QY 54 ATQTFATCINGVCWTVYHGAGTRTASPKGPVQMTYNDKDLVGQAPQGSRLTPCT 113
Db 1065 AAQTFATCINGVCWTVYHGAGTRTASPKGPVQMTYNDKDLVGQAPQGSRLTPCT 1124
QY 114 CGSSDLVLTTRHADVIPVRRGDSRGLSPRISYLYKSGSGGPLLCPAGHAGVIFRAAV 173
Db 1125 CGSSDLVLTTRHADVIPVRRGDSRGLSPRISYLYKSGSGGPLLCPAGHAGVIFRAAV 1184
QY 174 CTRGVAKAVDFIPVESLETTMRSP 197
Db 1185 CTRGVAKAVDFIPVENLETTMRSP 1208

RESULT 5
Q9ELS8 ID Q9ELS8 PRELIMINARY; PRT: 3011 AA.
AC Q9ELS8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=colonel;
RA Desai S.M., Devare S., Yamaguchi J.;
RT "Hepatitis C Virus."
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: AF290978; AAG02099.1; -.
DR HSSP: P27958; 1HEI.
DR InterPro: IPR000345; CytC_heme_bind.
DR InterPro: IPR001410; DEAD.

```



```

Query Match      85.6%; Score 883.5; DB 12; Length 3011;
Best Local Similarity 84.3%; Pred. No. 1.7e-79;
Matches 172; Conservative 8; Mismatches 15; Indels 9; Gaps 1;

QY 3 KGSVVIVGRIN-----LSGOTAYAAQOTRGEGCOETSGTGRDKNQVEGEVQIVST 53
DB 1005 RKGREILLGPADGWSKGRLLAPITAYAAQOTRGLGCIITSLTGRDKNQVEGEVQIVST 1064

QY 54 ATOTFLATCINGVCWTVYHGAGTGTIASPKGPVTOYTNVDKDLVGMQAPGSRSLTPCT 113
DB 1065 AATOTFLATCINGVCWTVYHGAGTGTIASPKGPVTOYTNVDKDLVGMQAPGSRSLTPCT 1124

QY 114 CGSSDLYLVTRHADVIPVRRGRDSRGLSPRISYLGKSSGGPLLCPCAGHAVGIFRAAV 173
DB 1125 CGSSDLYLVTRHADVIPVRRGRDSRGLSPRISYLGKSSGGPLLCPCAGHAVGIFRAAV 1184

QY 174 CTRGVAKAVDFIPVESLETTMRSP 197
DB 1185 CTRGVAKAVDFIPVESLETTMRSP 1208

RESULT 7
O36608 PRELIMINARY; PRT: 3011 AA.
AC O36608;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus strain H77.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=63746;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H77;
RX MEDLINE=97385173; PubMed=9238047;
RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;
RT "Transcripts from a single full-length cDNA clone of hepatitis C virus
RT are infectious when directly transfected into the liver of a
RT chimpanzee."
RL Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743(1997).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DB EMBL; AF011751; AAB67036.1; -
DB HSP; P27958; 1HEI.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR002517; HCV_NS3.
DR InterPro; IPR004109; HCV_NS4a.
DR InterPro; IPR000745; HCV_NS4b.
DR InterPro; IPR001490; HCV_NS4c.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PS_vir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; helicase_C; 1.

DR Pfam; PF00998; Viral_RdRP; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDc; 1.
DR PROSITE; PS50507; RDRP_POSITIVE; 1.
DR PROSITE; PS50521; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA; 327112 MW; 0B75E6B81CB5C198 CRC64;

Query Match      85.4%; Score 881.5; DB 12; Length 3011;
Best Local Similarity 83.8%; Pred. No. 2.7e-79;
Matches 171; Conservative 10; Mismatches 14; Indels 9; Gaps 1;

QY 3 KGSVVIVGRIN-----LSGDTAYAAQOTRGEGCOETSGTGRDKNQVEGEVQIVST 53
DB 1005 RRGQEILLGPADGWSKGRLLAPITAYAAQOTRGLGCIITSLTGRDKNQVEGEVQIVST 1064

QY 54 ATOTFLATCINGVCWTVYHGAGTGTIASPKGPVTOYTNVDKDLVGMQAPGSRSLTPCT 113
DB 1065 ATOTFLATCINGVCWTVYHGAGTGTIASPKGPVTOYTNVDKDLVGMQAPGSRSLTPCT 1124

QY 114 CGSSDLYLVTRHADVIPVRRGRDSRGLSPRISYLGKSSGGPLLCPCAGHAVGIFRAAV 173
DB 1125 CGSSDLYLVTRHADVIPVRRGRDSRGLSPRISYLGKSSGGPLLCPCAGHAVGIFRAAV 1184

QY 174 CTRGVAKAVDFIPVESLETTMRSP 197
DB 1185 CTRGVAKAVDFIPVESLETTMRSP 1208

RESULT 8
O9PWK5 PRELIMINARY; PRT: 3015 AA.
AC O9PWK5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99420396; PubMed=10489358;
RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;
RT "Hepatitis C virus: an infectious molecular clone of a second major
RT genotype (2a) and lack of viability of intertypic 1a and 2a
RT chimeras."
RL Virology 262:250-263(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Bukh J.;
RC Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DB EMBL; AF177040; AAF01182.1; -
DB EMBL; AF177038; AAF01180.1; -
DB HSP; P27958; 1HEI.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.

```



```

DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR002129; Pyridoxal_dec.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01539; HCV_core; 1.
DR Pfam: PF01560; HCV_NSI; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RDRP; 1.
DR Pfam: PF0186062; HCV_NSI; 1.
DR Pfam: PF0186062; HCV_NSI; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS00392; DDC_GAD_HDC_YDC; 1.
DR PROSITE: PS00507; RDRP_POSITIVE; 1.
DR PROSITE: PS0521; RDRP_VIRAL; 1.
DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3015 AA; 328159 MW; 87D3BC1F190663A CRC64;

Query Match 85.4%; Score 881.5; DB 12; Length 3015;
Best Local Similarity 83.8%; Pred. No. 2.7e-79;
Matches 171; Conservative 10; Mismatches 14; Indels 9; Gaps 1;

QY 3 KGSVVIVGRIN-----LSGDTAYAAQTRGEGCOETSGTGRDNQVEGEVQIVST 53
DB 1009 RRGQELLGPADGMVSKGWRLLAPITAYAAQTRGLGCIITSLTGRDNQVEGEVQIVST 1068

QY 54 ATOTFLATCINGVCWTVYHGAGTRTIASPKGPVTOMYTNVDKDLVGMQAPQGSRLTPTCT 113
DB 1069 ATOTFLATCINGVCWTVYHGAGTRTIASPKGPVIQMYTNVDQDLVGMQAPQGSRLTPTCT 1128

QY 114 CGSSDLYLVTRHADVIPVRRGDSRGLSPRPISYLGSSGGPLLCPCPAGHAGVIFRAAV 173
DB 1129 CGSSDLYLVTRHADVIPVRRGDSRGLSPRPISYLGSSGGPLLCPCPAGHAGVIFRAAV 1188

QY 174 CTRGVAKAVDFIPVESLETTMRSP 197
DB 1189 CTRGVAKAVDFIPVENLGTTRMRSP 1212

RESULT 9
Q9PMW9 PRELIMINARY; PRT; 3015 AA.
AC Q9PMW9
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
DE Hepatitis C virus.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OC NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-99420396; PubMed-10489358;
RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;
RT *Hepatitis C virus: an infectious molecular clone of a second major
RT genotype (2a) and lack of viability of intertypic 1a and 2a
RT chimeras.;
RL Virology 262:250-263(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Bukh J.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

```

```

CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: AF177039; AAF01181.1; -.
DR EMBL: AF177037; AAF01179.1; -.
DR HSSP: P27958; 1HEI.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NSI.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RDRP.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR002129; Pyridoxal_dec.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NSI; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RDRP; 1.
DR Pfam: PF0186062; HCV_NSI; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS00392; DDC_GAD_HDC_YDC; 1.
DR PROSITE: PS00507; RDRP_POSITIVE; 1.
DR PROSITE: PS0521; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3015 AA; 328084 MW; E309F6318067D6CD CRC64;

Query Match 85.4%; Score 881.5; DB 12; Length 3015;
Best Local Similarity 83.8%; Pred. No. 2.7e-79;
Matches 171; Conservative 10; Mismatches 14; Indels 9; Gaps 1;

QY 3 KGSVVIVGRIN-----LSGDTAYAAQTRGEGCOETSGTGRDNQVEGEVQIVST 53
DB 1009 RRGQELLGPADGMVSKGWRLLAPITAYAAQTRGLGCIITSLTGRDNQVEGEVQIVST 1068

QY 54 ATOTFLATCINGVCWTVYHGAGTRTIASPKGPVTOMYTNVDKDLVGMQAPQGSRLTPTCT 113
DB 1069 ATOTFLATCINGVCWTVYHGAGTRTIASPKGPVIQMYTNVDQDLVGMQAPQGSRLTPTCT 1128

QY 114 CGSSDLYLVTRHADVIPVRRGDSRGLSPRPISYLGSSGGPLLCPCPAGHAGVIFRAAV 173
DB 1129 CGSSDLYLVTRHADVIPVRRGDSRGLSPRPISYLGSSGGPLLCPCPAGHAGVIFRAAV 1188

QY 174 CTRGVAKAVDFIPVESLETTMRSP 197
DB 1189 CTRGVAKAVDFIPVENLGTTRMRSP 1212

RESULT 10
Q91RR8 PRELIMINARY; PRT; 181 AA.
AC Q91RR8
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NS3 protease (fragment).
DE Hepatitis C virus.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OC NCBI_TaxID=11103;

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 30, 2003, 19:18:33 ; Search time 2560.57 Seconds
(without alignments)
3147.423 Million cell updates/sec

Title: US-09-965-594-16

Perfect score: 1032

Sequence: 1 MKKKGSVIVGRINLSGDTA.....VAKAVDFIPVESLETHRSP 197

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-O=/cgn2_1/USPTO.spool/US0965594/runat_29082003_151919_28310/app_query.fasta_1.2872
-DB=GenEmbl -OPMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blast62 -TRANS=human40.cdi -LIST=45
-DOCLALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0965594.ecgn_1_14686 -runat_29082003_151919_28310 -NCPU=6 -ICPU=3
-NO_JMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*
1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vi:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.mu:*
20: em.om:*
21: em.or:*
22: em.ov:*
23: em.pat:*
24: em.ph:*
25: em.pl:*
26: em.ro:*
27: em.sts:*
28: em.un:*

29: em.vi:*
30: em.htg_hum:*
31: em.htg_inv:*
32: em.htg_other:*
33: em.htg_mus:*
34: em.htg_pin:*
35: em.htg_rtd:*
36: em.htg_mam:*
37: em.htg_vrt:*
38: em.sy:*
39: em.htgo_hum:*
40: em.htgo_mus:*
41: em.htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	922.5	89.4	12734	6	ARI79057 Sequence
2	911.5	88.3	1398	6	ARI45264 Sequence
3	908.5	88.0	1998	6	ARI45268 Sequence
4	907.5	87.9	1998	6	ARI45262 Sequence
5	907.5	87.9	1998	6	ARI45263 Sequence
6	904.5	87.6	651	6	ARI45254 Sequence
7	904.5	87.6	1998	6	ARI45266 Sequence
8	904.5	87.6	1998	6	ARI45267 Sequence
9	903.5	87.5	1998	6	ARI45261 Sequence
10	903.5	87.5	2016	6	ARI45269 Sequence
11	902.5	87.5	12734	14	AF268278 Pestivirus
12	901.5	87.4	651	6	ARI45258 Sequence
13	900.5	87.3	651	6	ARI45252 Sequence
14	900.5	87.3	651	6	ARI45253 Sequence
15	900.5	87.3	1998	6	ARI45265 Sequence
16	900.5	87.3	2016	6	ARI45270 Sequence
17	900	87.2	648	6	ARI45274 Sequence
18	898	87.0	648	6	ARI45272 Sequence
19	897.5	87.0	651	6	ARI45256 Sequence
20	897.5	87.0	651	6	ARI45257 Sequence
21	897.5	87.0	651	6	ARI45260 Sequence
22	896.5	86.9	651	6	ARI45251 Sequence
23	896	86.8	648	6	ARI45273 Sequence
24	894	86.6	648	6	ARI45271 Sequence
25	893.5	86.6	651	6	ARI45255 Sequence
26	893.5	86.6	651	6	ARI45259 Sequence
27	891	86.3	8157	6	ARI27810 Sequence
28	891	86.3	8157	6	BD081911 Hepatitis
29	889	86.1	1932	6	ARI27809 Sequence
30	889	86.1	1932	6	BD081910 Hepatitis
31	888.5	86.1	9646	6	ARI10828 Sequence
32	888.5	86.1	9646	6	BD069982 Functiona
33	888.5	86.1	9646	14	AF009606 Hepatitis
34	888.5	86.1	12980	6	ARI10831 Sequence
35	888.5	86.1	12980	6	BD069985 Functiona
36	884.5	85.7	5360	6	ARI18686 Sequence
37	884.5	85.7	5360	6	I06434 Sequence 48
38	884.5	85.7	5360	6	I09328 Sequence 8
39	884.5	85.7	6785	6	ARI18692 Sequence
40	884.5	85.7	6785	6	I06440 Sequence 54
41	884.5	85.7	6785	6	I09329 Sequence 10
42	884.5	85.7	7310	6	ARI18696 Sequence
43	884.5	85.7	7310	6	I09331 Sequence 15
44	884.5	85.7	7310	14	HPCPOLYP M32084 Hepatitis C
45	884.5	85.7	8316	6	ARI18703 Sequence

ALIGNMENTS

RESULT 1

AR179057
LOCUS AR179057 12734 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6326137.
ACCESSION AR179057
VERSION AR179057.1 GI:20220612
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 12734)
AUTHORS Hong Z., Lai V.C.H. and Lau J.Y.N.
TITLE Hepatitis C virus protease-dependent chimeric pestivirus
JOURNAL Patent: US 6326137-A 1 04-DEC-2001;
FEATURES
Location/Qualifiers
source 1..12734 /organism="unknown"
BASE COUNT 4032 a 2604 c 3295 g 2803 t
ORIGIN
Alignment Scores:
Pred. No.: 8,58e-67 Length: 12734
Score: 922.50 Matches: 180
Percent Similarity: 94.36% Conservative: 4
Best Local Similarity: 92.31% Mismatches: 8
Query Match: 89.39% Indels: 3
DB: 6 Gaps: 1

US-09-965-594-16 (1-197) x AR179057 (1-12734)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
DB 413 GGTAGTGTGTTATTTGTTAGTAATTTGTTTATCTCTGTTAGTGTAGTACACGGGGGTAC 472
QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGlnThrSerGlnThrGlyArgAspLys 41
DB 473 GCCCAGCAGCAGCAGCGGCTCTAGGGTGTAGAGTACCACTGTGACTGCGCGGGACAAA 532
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
DB 533 AACCAAGTGGAGGGTGAGTGCAGATCGTGTCAACTGCTACCCAAACCTTCCCTGGCAACG 592
QY 62 CysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
DB 593 TGCATCAATGGGGTATGCTGACTGTCTACACCGGGCGGACAGGAGCATCGCATCA 652
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101
DB 653 CCCAAGGCTCTGTCATCCAGATGTATACCAATGTGCACCAAGACCTTGTGGCTGGCCC 712
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
DB 713 GCTCTCAAGGTTCGGGTCTATTGACACCTGACACCTGCGGCTCTCGGACCTTTACCTG 772
QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyArgGlyAspSerArgGlySerLeu 141
DB 773 GTTAGCAGCAGCAGCGAGCTCATCTCCGTGCGCGGAGGTGATACAGGGGTAGCTG 832
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerGlyGlyProLeuLeuCysPro 161
DB 833 CTTTCGCCCGGCCCATTTCTACCTAAAGGCTCCCTCGGGGGTCCGCTGTTGTGCCCC 892
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181
DB 893 GCGGACAGCAGCGGTGGGCTATTACGGCCCGGTGTGACCCCGTGGAGTGGCCAGGCG 952
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196
DB 953 GTGGACTTATCCCTGTGGAGAACCTAGACAGAACCATGAGATCC 997

RESULT 2
LOCUS AR145264
DEFINITION Sequence 105 from patent US 6211338.
AR145264
ACCESSION AR145268
VERSION AR145268.1 GI:15107135

AR145264
VERSION AR145264.1 GI:15107131
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1998)
AUTHORS Malcolm, B.A., Taremi, S. Shane., Weber, P.C. and Yao, N.
TITLE Single-chain recombinant complexes of hepatitis C virus NS3
protease and NS4A cofactor peptide
JOURNAL Patent: US 6211338-A 105 03-APR-2001;
FEATURES
Location/Qualifiers
source 1..1998 /organism="unknown"
BASE COUNT 411 a 595 c 569 g 423 t
ORIGIN
Alignment Scores:
Pred. No.: 9,21e-67 Length: 1998
Score: 911.50 Matches: 170
Percent Similarity: 94.90% Conservative: 16
Best Local Similarity: 86.73% Mismatches: 7
Query Match: 88.32% Indels: 3
DB: 6 Gaps: 1

US-09-965-594-16 (1-197) x AR145264 (1-1998)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
DB 64 GGTCTGTGTATTGTTGTGTAGTAATTTATTTACTGTGTAGTGTATACAGCGCTAC 123
QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGlnThrSerGlnThrGlyArgAspLys 41
DB 124 TCCCAACAGACGCGGGGCTACTTGGTTGCAAGAGACTAGCCTTACAGCGCGGACAAG 183
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
DB 184 AACCAAGTGGAGGGTGGAGGTTCAGGTGTCTCCACCGCAACAAATCTTCTCTGGGACC 243
QY 62 CysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
DB 244 TCGGTCAACGGGTGTGTGGACGTTTACCATGTGCTGGCTCAAGACCTTAGCGGC 303
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101
DB 304 CCAAGGGGCGCAATCACCCAGATGTACACTAATGTGACAGGACCTCGTGGCTGGCAG 363
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
DB 364 GCGCCCGCGGGCGGCTCTCTTGACACCATGACCTGTGGCAGCTCAGACCTTACTTG 423
QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyArgGlyAspSerArgGlySerLeu 141
DB 424 GTCAGAGACATGCTGACGTCTATCCGGTGGCGCGGCGGAGACAGTAGGGGAGCCTG 483
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerGlyGlyProLeuLeuCysPro 161
DB 484 CTTCTCCCGAGCGCTGTCTCTACTTGAAGGGCTCTCCGGGTGTGCCACTGCTGCTGCCCT 543
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181
DB 544 TCGGGGACGCTGTGGGCTCTTCGGGCTGCCGTATGCACCCCGGGGGTTCGGAAGCG 603
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
DB 604 GTGGACTTGTGCGGTAGTCCATGGAAACTACTATGCGGTCTCG 651

RESULT 3
LOCUS AR145268
DEFINITION Sequence 109 from patent US 6211338.
AR145268
ACCESSION AR145268
VERSION AR145268.1 GI:15107135

KEYWORDS

SOURCE Unknown.
ORGANISM Unknown.

REFERENCE

1 (bases 1 to 1998)
AUTHORS Malcolm,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.
TITLE Single-chain recombinant complexes of hepatitis C virus NS3
protease and NS4A cofactor peptide
JOURNAL Patent: US 6211338-A 109 03-APR-2001;
FEATURES Location/Qualifiers
1..1998
/organism="unknown"

BASE COUNT

411 a 595 c 569 g 423 t
ORIGIN

Alignment Scores:

Pred. No.: 1.64e-66 Length: 1998
Score: 908.50 Matches: 169
Percent Similarity: 94.90% Conservative: 17
Best Local Similarity: 86.22% Mismatches: 7
Query Match: 88.03% Indels: 3
DB: 6 Gaps: 1

US-09-965-594-16 (1-197) x AR145268 (1-1998)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
|||||
DB 64 GGTCTGTTGTTATTTGTTGTTAGAAATTATTTATCTGTTAGTAGTATCATCGGCTAC 123
QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41
|||||
DB 124 TCCCAACAGACGCGGGGCTACTTTGGTTGCAAGAGATCACTAGCCTTACAGCCGGGCAAG 183
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
|||||
DB 184 ACCAGGTCGAGGAGAGGTTTCAGGTGTTTCCACCGCAACAACTCTTCTGGCGACC 243
QY 62 CysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
|||||
DB 244 TCGCTCAACGCGGTGTTGACCGCTTACCATGCTGCTGCTCAAGACCTTAGCGCGC 303
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101
DB 304 CCAAGGGGCCAATCACCATGTACACTAATGTGGACGAGACCTCGTGGCTGGCAG 363
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
DB 364 GCGCCCGCGGGCGGTCTCTTGCACCATGCACCTGTCGACCTTACTTG 423
QY 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141
DB 424 CTCACGACACATGCTGACGTCTATTCGGGTGCGCGGGCGGCGACAGTAGGGGAGCCTG 483
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
DB 484 CTCCTCCCGCAGCTCTCTCTACTTGAAGGCTCTGCTGCTCCACTCTGCGCT 543
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181
DB 544 TCGGGGCACTGTGGGATCTTCGGGCTCCGCTATGCACCGGGGGGTTGCAAGGCG 603
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
DB 604 GTGGACTTTGTGCCGTAGAGTCCATGGAACTACTATGCGGTCTCCG 651

RESULT 4

AR145262
LOCUS AR145262 1998 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 103 from patent US 6211338.
ACCESSION AR145262
VERSION AR145262.1 GI:15107129
KEYWORDS
SOURCE Unknown.

ORGANISM

Unknown.
Unclassified.

REFERENCE

1 (bases 1 to 1998)
AUTHORS Malcolm,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.
TITLE Single-chain recombinant complexes of hepatitis C virus NS3
protease and NS4A cofactor peptide
JOURNAL Patent: US 6211338-A 103 03-APR-2001;
FEATURES Location/Qualifiers
1..1998
/organism="unknown"

BASE COUNT

410 a 596 c 568 g 424 t
ORIGIN

Alignment Scores:

Pred. No.: 1.98e-66 Length: 1998
Score: 907.50 Matches: 170
Percent Similarity: 94.39% Conservative: 15
Best Local Similarity: 86.73% Mismatches: 8
Query Match: 87.94% Indels: 3
DB: 6 Gaps: 1

US-09-965-594-16 (1-197) x AR145262 (1-1998)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
|||||
DB 64 GGTCTGTTGTTATTTGTTGTTAGAAATTATTTATCTGTTAGTAGTATCATCGGCTAC 123
QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41
|||||
DB 124 TCCCAACAGACGCGGGGCTACTTTGGTTGCAAGATCACTAGCCTTACAGCCGGGCAAG 183
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
|||||
DB 184 ACCAGGTCGAGGAGAGGTTTCAGGTGTTTCCACCGCAACAACTCTTCTGGCGACC 243
QY 62 CysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
|||||
DB 244 TCGCTCAACGCGGTGTTGACCGCTTACCATGCTGCTGCTCAAGACCTTAGCGCGC 303
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101
DB 304 CCAAGGGGCCAATCACCATGTACACTAATGTGGACGAGACCTCGTGGCTGGCAG 363
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
DB 364 GCGCCCGCGGGCGGTCTCTTGCACCATGCACCTGTCGACCTTACTTG 423
QY 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141
DB 424 CTCACGACACATGCTGACGTCTATTCGGGTGCGCGGGCGGCGACAGTAGGGGAGCCTG 483
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
DB 484 CTCCTCCCGCAGCTCTCTCTACTTGAAGGCTCTGCTGCTCCACTCTGCGCT 543
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181
DB 544 TCGGGGCACTGTGGGATCTTCGGGCTCCGCTATGCACCGGGGGGTTGCAAGGCG 603
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
DB 604 GTGGACTTTGTGCCGTAGAGTCCATGGAACTACTATGCGGTCTCCG 651

RESULT 5

AR145263
LOCUS AR145263 1998 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 104 from patent US 6211338.
ACCESSION AR145263
VERSION AR145263.1 GI:15107130
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.

REFERENCE 1 (bases 1 to 1998)
 AUTHORS Malcolm, B.A., Taremi, S. Shane., Weber, P.C. and Yao, N.
 TITLE Single-chain recombinant complexes of hepatitis C virus NS3
 protease and NS4A cofactor peptide
 Patent: US 6211338-A 104 03-APR-2001;
 JOURNAL Location/Qualifiers
 FEATURES source 1. .1998
 BASE COUNT 410 a 596 c 568 g 424 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1.98e-66 Length: 1998
 Score: 907.50 Matches: 170
 Percent Similarity: 94.39% Conservative: 15
 Best Local Similarity: 86.73% Mismatches: 8
 Query Match: 87.94% Indels: 3
 DB: 6

US-09-965-594-16 (1-197) x ARI45263 (1-1998)

Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
 |||||
 Db 64 GGTCTGTTGTTATTGTTGGTAGAATTATTTATCTGGTAGTGGTAGTATCAGGCGCTAC 123
 Qy 22 AlaGlnGlnThrArgGlyGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41
 |||||
 Db 124 TCCCAACAGACGCGGGCGCTACTTGGTTCATCAGACTAGCTTACAGCGCGGCAAG 183
 Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
 |||||
 Db 184 AACGAGTCCAGGAGAGGTTTCAGGTGGTTCCACCGCAACAATCTCTCTCGCGACC 243
 Qy 62 CysIleAsnGlyValCysIleThrValThrHisGlyAlaGlyThrArgThrIleAlaSer 81
 |||||
 Db 244 TGGTCAACGGCGTGTGGACCGTTTACCATGGTGGTGGTCAAGACCTTAGCGGC 303
 Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
 |||||
 Db 304 CCAAGGGCGCAATCACCAGATGTACACTAATGTGGACCAAGACCTTCGTGGCTGGCAG 363
 Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
 |||||
 Db 364 GCGCGCCCGCGGGCGGCTTCTTGCACACCATGCACCTGTGGCAGCTCAGACCTTTACTTG 423
 Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
 |||||
 Db 424 GTCCAGAGACATCTGACGTCATTCGGTGGCGCGCGGCGGCGACACTAGGGGAGCGCTG 483
 Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
 |||||
 Db 484 CTCTCCCGCAGGCTGTCTCTACTTGAAGGGCTCTTCGGGTGGTCCACTGCTCGCCCT 543
 Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181
 |||||
 Db 544 TCGGGGACGCTGTGGCATCTTCGGGTGGCGGTATGCACCGGGGGTTCGGAAGCG 603
 Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
 |||||
 Db 604 GTGGACTTTGTGCCCGTAGAGTCCATGGAAACTACTATGCGGTCTCCG 651

RESULT 6
 ARI45254
 LOCUS ARI45254 651 bp DNA linear PAT 08-AUG-2001
 DEFINITION Sequence 95 from patent US 6211338.
 ACCESSION ARI45254
 VERSION ARI45254.1 GI:15107121
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 651)
 AUTHORS Malcolm, B.A., Taremi, S. Shane., Weber, P.C. and Yao, N.

TITLE Single-chain recombinant complexes of hepatitis C virus NS3
 protease and NS4A cofactor peptide
 Patent: US 6211338-A 95 03-APR-2001;
 JOURNAL Location/Qualifiers
 FEATURES source 1. .651
 BASE COUNT 120 a 187 c 200 g 144 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1.03e-66 Length: 651
 Score: 904.50 Matches: 169
 Percent Similarity: 94.87% Conservative: 16
 Best Local Similarity: 86.67% Mismatches: 7
 Query Match: 87.65% Indels: 3
 DB: 6

US-09-965-594-16 (1-197) x ARI45254 (1-651)

Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
 |||||
 Db 64 GGTCTGTTGTTATTGTTGGTAGAATTATTTATCTGGTAGTGGTAGTATCAGGCGCTAC 123
 Qy 22 AlaGlnGlnThrArgGlyGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41
 |||||
 Db 124 TCCCAACAGACGCGGGCGCTACTTGGTTCGCAAGACTAGCTTACAGCGCGGCAAG 183
 Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
 |||||
 Db 184 AACGAGTCCAGGAGAGGTTTCAGGTGGTTTCACCGCAACAATCTCTCTCGCGACC 243
 Qy 62 CysIleAsnGlyValCysIleThrValThrHisGlyAlaGlyThrArgThrIleAlaSer 81
 |||||
 Db 244 TGGTCAACGGCGTGTGGACCGTTTACCATGGTGGTGGTCAAGACCTTAGCGGC 303
 Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
 |||||
 Db 304 CCAAGGGCGCAATCACCAGATGTACACTAATGTGGACCAAGACCTTCGTGGCTGGCAG 363
 Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
 |||||
 Db 364 GCGCGCCCGCGGGCGGCTTCTTGCACACCATGCACCTGTGGCAGCTCAGACCTTTACTTG 423
 Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
 |||||
 Db 424 GTCCAGAGACATCTGACGTCATTCGGTGGCGCGGCGGCGGCGACACTAGGGGAGCGCTG 483
 Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
 |||||
 Db 484 CTCTCCCGCAGGCTGTCTCTACTTGAAGGGCTCTTCGGGTGGTCCACTGCTCGCCCT 543
 Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181
 |||||
 Db 544 TCGGGGACGCTGTGGCATCTTCGGGTGGCGGTATGCACCGGGGGTTCGGAAGCG 603
 Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196
 |||||
 Db 604 GTGGACTTTGTGCCCGTAGAGTCCATGGAAACTACTATGCGGTCT 648

RESULT 7
 ARI45266
 LOCUS ARI45266 1998 bp DNA linear PAT 08-AUG-2001
 DEFINITION Sequence 107 from patent US 6211338.
 ACCESSION ARI45266
 VERSION ARI45266.1 GI:15107133
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 1998)
 AUTHORS Malcolm, B.A., Taremi, S. Shane., Weber, P.C. and Yao, N.
 TITLE Single-chain recombinant complexes of hepatitis C virus NS3
 protease and NS4A cofactor peptide

JOURNAL Patent: US 6211338-A 107 03-APR-2001:
FEATURES Location/Qualifiers

source 1..1998 /organism="unknown"
BASE COUNT 410 a 596 c 568 g 424 t
ORIGIN

Alignment Scores:
Pred. No.: 3,53e-66 Length: 1998
Score: 904.50 Matches: 169
Percent Similarity: 94.39% Conservative: 16
Best Local Similarity: 86.22% Mismatches: 8
Query Match: 87.65% Indels: 3
DB: 6 Gaps: 1

US-09-965-594-16 (1-197) x ARL45266 (1-1998)

```
QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
|||||
Db 64 GGTTCGTGTTATTGTTGTTAGAAATATTATTATCTGTTAGTATCATCAGGCTAC 123
QY 22 AlaGlnGlnThrArgGlyGluGlyCysGlnGlnThrSerGlnThrGlyArgAspLys 41
|||||
Db 124 TCCCAACACAGCGCGGCTACTTGGTTGCAAGATCACTAGCCTTACAGCGCGGACAAG 183
QY 42 AsnGlnValGluGlyGlnValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
|||||
Db 184 AACCAAGTTCGAGGAGAGGTTCCAGTGGTTTCCACCGCAACACAATCTCTCTGGCGACC 243
QY 62 CysIleAsnGlyValCysTrpThrValThrHisGlyAlaGlyThrArgThrIleAlaSer 81
|||||
Db 244 TCGCTCAACGCGGTGTGTGACCGTTTACCATGGTGTGGCTCAAAAGACCTTAGCCGCG 303
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
|||||
Db 304 CCNAGGGGCGCAATCACCAGATGATACACTAATGTGGACCAAGACCTCGCGGCTGGCAG 363
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
|||||
Db 364 CGCGCCCGCGGCGGCTTCTTGCACCATGCACCTGTGCAGCTCAGACCTTTACTTG 423
QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
|||||
Db 424 GTCCAGAGACATGTGAGTCTTCCCTACTTGAAGGCTCTGCTGGTCCACTGCTGCGCT 543
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
|||||
Db 484 CTCCTCCCGACGCTCTCTCTACTTGAAGGCTCTGCTGGTCCACTGCTGCGCT 543
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181
|||||
Db 544 TCGGGGACGCTGTGGGATCTTCCGGCTGCCGTATGCACCGCGGGGTTGCGAAGCG 603
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
|||||
Db 604 GTGGACTTTGTGCGCGTAGAGTCCATGGAAACTACTATGCGGTCTCCG 651
```

RESULT 8
ARL45267
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES

1 (bases 1 to 1998)
Malcolm,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.
Single-chain recombinant complexes of hepatitis C virus NS3
protease and NS4A cofactor peptide
Patent: US 6211338-A 108 03-APR-2001;
Location/Qualifiers

source 1..1998 /organism="unknown"

source 1..1998 /organism="unknown"

BASE COUNT 410 a 596 c 568 g 424 t
ORIGIN

Alignment Scores:
Pred. No.: 3,53e-66 Length: 1998
Score: 904.50 Matches: 169
Percent Similarity: 94.39% Conservative: 16
Best Local Similarity: 86.22% Mismatches: 8
Query Match: 87.65% Indels: 3
DB: 6 Gaps: 1

US-09-965-594-16 (1-197) x ARL45267 (1-1998)

```
QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
|||||
Db 64 GGTTCGTGTTATTGTTGTTAGAAATATTATTATCTGTTAGTATCATCAGGCTAC 123
QY 22 AlaGlnGlnThrArgGlyGluGlyCysGlnGlnThrSerGlnThrGlyArgAspLys 41
|||||
Db 124 TCCCAACACAGCGCGGCTACTTGGTTGCAAGATCACTAGCCTTACAGCGCGGACAAG 183
QY 42 AsnGlnValGluGlyGlnValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
|||||
Db 184 AACCAAGTTCGAGGAGAGGTTCCAGTGGTTTCCACCGCAACACAATCTCTCTGGCGACC 243
QY 62 CysIleAsnGlyValCysTrpThrValThrHisGlyAlaGlyThrArgThrIleAlaSer 81
|||||
Db 244 TCGCTCAACGCGGTGTGTGACCGTTTACCATGGTGTGGCTCAAAAGACCTTAGCCGCG 303
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
|||||
Db 304 CCNAGGGGCGCAATCACCAGATGATACACTAATGTGGACCAAGACCTCGCGGCTGGCAG 363
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
|||||
Db 364 CGCGCCCGCGGCGGCTTCTTGCACCATGCACCTGTGCAGCTCAGACCTTTACTTG 423
QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
|||||
Db 424 GTCCAGAGACATGTGAGTCTTCCCTACTTGAAGGCTCTGCTGGTCCACTGCTGCGCT 543
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
|||||
Db 484 CTCCTCCCGACGCTCTCTCTACTTGAAGGCTCTGCTGGTCCACTGCTGCGCT 543
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181
|||||
Db 544 TCGGGGACGCTGTGGGATCTTCCGGCTGCCGTATGCACCGCGGGGTTGCGAAGCG 603
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
|||||
Db 604 GTGGACTTTGTGCGCGTAGAGTCCATGGAAACTACTATGCGGTCTCCG 651
```

RESULT 9
ARL45261
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES

1 (bases 1 to 1998)
Malcolm,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.
Single-chain recombinant complexes of hepatitis C virus NS3
protease and NS4A cofactor peptide
Patent: US 6211338-A 102 03-APR-2001;
Location/Qualifiers

source 1..1998 /organism="unknown"

BASE COUNT 409 a 597 c 567 g 425 t
ORIGIN

Alignment Scores:

Pred. No.: 4,28e-66 Length: 1998
Score: 903.50 Matches: 170
Percent Similarity: 93.88% Conservative: 14
Best Local Similarity: 86.73% Mismatches: 9
Query Match: 87.55% Indels: 3
DB: 6 Gaps: 1

US-09-965-594-16 (1-197) x AR145261 (1-1998)

QY 5 GlySerValValIleValGlyArgGileAsnLeuSerGlyAsp-----ThralaTyr 21
|||||
DB 64 GGTTCTGTTGTTATTTGTTAGAAATTTATTTATCTGGTAGTGTAGTATCATCGGCGCTAC 123
QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGlnThrSerGlnThrGlyArgAspLys 41
|||||
DB 124 TCCCAACAGACGGCGGCGCTACTTGGTTGCATCATCACTAGCCTTACAGCGCGGGAACAG 183
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
|||||
DB 184 AACAGGTCAGGAGAGGTTTCAGTGGTTTCCACCGCAACAAATCCTTCTCGGCGACC 243
QY 62 CysIleAsnGlyValCysTrrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
|||||
DB 244 TGCCTCAACGGCGGTGTGTGGCGGTTTACATGGTGTGGCTCAAGAGACCTTAGCGCGC 303
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrrPglN 101
|||||
DB 304 CCAAGGGGCAATCACCAAGATGTACATTAATGTGGACCGACCTCGTCGCGCTGGCAG 363
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
|||||
DB 364 GCGCCCCCGGGCGGCTTCTTGCACACCATGCACCTGTGGCAGCTCAGACCTTTACTTG 423
QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
|||||
DB 424 GTCACGAGACATGCTGACGCTATTCGGTGGCGCGCGGCGGCGACAGTAGGGGAGCGCTG 483
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
|||||
DB 484 CTCTCCCCAGGCGTCTCTACTTGAAGGGCTCTCGGGTGTCTCCACTGCTCTGCCCT 543
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181
|||||
DB 544 TCGGGGACGCTGTGGGCATCTTCCGGGCTGCGGTATGCACCGGGGGGTTCGGAAGCG 603
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
|||||
DB 604 GTGGACTTTGGCCCGTAGAGTCCATGGAACACTACTATGCGGCTCTCG 651

RESULT 10

AR145269
LOCUS AR145269 2016 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 110 from patent US 6211338.
ACCESSION AR145269
VERSION AR145269.1 GI:15107136

KEYWORDS

Unknown.

ORGANISM

Unknown.

REFERENCE

1 (bases 1 to 2016)

Malcolm, B.A., Taremi, S., Shane, J., Weber, P.C. and Yao, N.

Single-chain recombinant complexes of hepatitis C virus NS3

protease and NS4A cofactor peptide

Patent: US 6211338-A 110 03-APR-2001;

Location/Qualifiers

1..2016

/organism="unknown"

BASE COUNT 412 a 603 c 570 g 431 t

ORIGIN

Alignment Scores:

Pred. No.: 4,32e-66 Length: 2016
Score: 903.50 Matches: 170
Percent Similarity: 93.88% Conservative: 14
Best Local Similarity: 86.73% Mismatches: 9
Query Match: 87.55% Indels: 3
DB: 6 Gaps: 1

US-09-965-594-16 (1-197) x AR145269 (1-2016)

QY 5 GlySerValValIleValGlyArgGileAsnLeuSerGlyAsp-----ThralaTyr 21
|||||
DB 82 GGTTCTGTTGTTATTTGTTAGAAATTTATTTATCTGGTAGTGTAGTATCATCGGCGCTAC 141
QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGlnThrSerGlnThrGlyArgAspLys 41
|||||
DB 142 TCCCAACAGACGGCGGCGCTACTTGGTTGCATCATCACTAGCCTTACAGCGCGGGAACAG 201
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
|||||
DB 202 AACAGGTCAGGAGAGGTTTCAGTGGTTTCCACCGCAACAAATCCTTCTCGGCGACC 261
QY 62 CysIleAsnGlyValCysTrrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
|||||
DB 262 TGCCTCAACGGCGGTGTGTGGCGGTTTACATGGTGTGGCTCAAGAGACCTTAGCGCGC 321
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrrPglN 101
|||||
DB 322 CCAAGGGGCAATCACCAAGATGTACATTAATGTGGACCGACCTCGTCGCGCTGGCAG 381
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
|||||
DB 382 GCGCCCCCGGGCGGCTTCTTGCACACCATGCACCTGTGGCAGCTCAGACCTTTACTTG 441
QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
|||||
DB 442 GTCACGAGACATGCTGACGCTATTCGGTGGCGCGGCGGCGACAGTAGGGGAGCGCTG 501
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
|||||
DB 502 CTCTCCCCAGGCGTCTCTACTTGAAGGGCTCTCGGGTGTCTCCACTGCTCTGCCCT 561
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181
|||||
DB 562 TCGGGGACGCTGTGGGCATCTTCCGGGCTGCGGTATGCACCGGGGGGTTCGGAAGCG 621
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
|||||
DB 622 GTGGACTTTGGCCCGTAGAGTCCATGGAACACTACTATGCGGCTCTCG 669

RESULT 11

AF268278
LOCUS AF268278 12734 bp RNA linear VRL 12-JUL-2000
DEFINITION Pestivirus type 1, complete genome.
ACCESSION AF268278
VERSION AF268278.1 GI:9049956

KEYWORDS

Pestivirus type 1

Pestivirus type 1

Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

Pestivirus.

1 (bases 1 to 12734)

Lai, V.C., Zhong, W., Skelton, A., Ingravallo, P., Vassiliev, V.,

Donis, R.O., Hong, Z. and Lau, J.Y.

Generation and characterization of a hepatitis C virus NS3

protease-dependent bovine viral diarrhea virus

J. Virol. 74 (14), 6339-6347 (2000)

20323484

10864644

2 (bases 1 to 12734)

Lai, V.C.H. and Hong, Z.

Direct Submission

Unclassified.

REFERENCE 1 (bases 1 to 651)
 AUTHORS Malcolm,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.
 TITLE Single-chain recombinant complexes of hepatitis C virus NS3
 JOURNAL protease and NS4A cofactor peptide
 PATENT: US 6211338-A 99 03-APR-2001;
 FEATURES Location/Qualifiers
 source 1. .651

BASE COUNT 120 a 187 c 200 g 144 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1.83e-66 Length: 651
 Score: 901.50 Matches: 168
 Percent Similarity: 94.87% Conservative: 17
 Best Local Similarity: 86.15% Mismatches: 7
 Query Match: 87.35% Indels: 3
 DB: 6 Gaps: 1

US-09-965-594-16 (1-197) x AR145258 (1-651)

QY 5 GlySerValValIleValGlyValArgIleAsnLeuSerGlyGlyAsp-----ThrAlaTyr 21
 DB 64 GGTTCCTGTTATTTGGTTAGTAATTTATTCGTAGTGGTAGTATCATCGGGCTAC 123
 QY 22 AlaGlnGlnThrArgGlyGluGlyCysGlnGlnThrSerGlnThrGlyArgAspLys 41
 DB 124 TCCCAACAGACGGGGCCTACTTGGTTGCAAGACATAGCCTTACAGCGGGACAAG 183
 QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
 DB 184 AACAGATCGAGGAGAGGTTCAGGTGGTTCCACCGCAACACAATCCCTTCCTGGCGACC 243
 QY 62 CysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
 DB 244 TGGCTCAACGGCGTGTGTGGACCGTTTACCATGGTGTGCTGCTCAAGACCTTAGCGGC 303
 QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
 DB 304 CCAAGGGGCGCAATCACCCAGATGTACACTAATGTGGACGAGGACCTCGTCGCTGGCAG 363
 QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
 DB 364 CGCGCCCGGGGGCGGTCTCTTGACACCATGACCTGTGGCAGCTCAGACCTTTACTTG 423
 QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
 DB 424 CTCACGAGACATGCTGACGTCATTCCGGTGGCGGGGGCGGACAGTAGGGGGAGCCTG 483
 QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerGlyGlyProLeuLeuCysPro 161
 DB 484 CTCCTCCCGCGGCGTCTCTCTACTTGAAGGGCTCTCGGGTGGTCCACTGCTCGCCT 543
 QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181
 DB 544 TCGGGGACGCTGTGGGCATCTCCGGCTGCGGCTGACCCCGGGGGTTGCAAGGCG 603
 QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196
 DB 604 GTGGACTTTGTGCGCGTAGATGCTCATGGAACACTACTATGCGGTCT 648

RESULT 13
 LOCUS AR145252 651 bp DNA linear PAT 08-AUG-2001
 DEFINITION Sequence 93 from patent US 6211338.
 ACCESSION AR145252
 VERSION AR145252.1 GI:15107119
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 651)

AUTHORS Malcolm,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.
 TITLE Single-chain recombinant complexes of hepatitis C virus NS3
 JOURNAL protease and NS4A cofactor peptide
 PATENT: US 6211338-A 93 03-APR-2001;
 FEATURES Location/Qualifiers
 source 1. .651

BASE COUNT 119 a 188 c 199 g 145 t
 ORIGIN

Alignment Scores:
 Pred. No.: 2.21e-66 Length: 651
 Score: 900.50 Matches: 169
 Percent Similarity: 94.36% Conservative: 15
 Best Local Similarity: 86.67% Mismatches: 8
 Query Match: 87.26% Indels: 3
 DB: 6 Gaps: 1

US-09-965-594-16 (1-197) x AR145252 (1-651)

QY 5 GlySerValValIleValGlyValArgIleAsnLeuSerGlyGlyAsp-----ThrAlaTyr 21
 DB 64 GGTTCCTGTTATTTGGTTAGTAATTTATTCGTAGTGGTAGTATCATCGGGCTAC 123
 QY 22 AlaGlnGlnThrArgGlyGluGlyCysGlnGlnThrSerGlnThrGlyArgAspLys 41
 DB 124 TCCCAACAGACGGGGCCTACTTGGTTGCAAGATCAGCCTTACAGCGGGACAAG 183
 QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
 DB 184 AACAGATCGAGGAGAGGTTCAGGTGGTTCCACCGCAACACAATCCCTTCCTGGCGACC 243
 QY 62 CysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
 DB 244 TGGCTCAACGGCGTGTGTGGACCGTTTACCATGGTGTGCTGCTCAAGACCTTAGCGGC 303
 QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
 DB 304 CCAAGGGGCGCAATCACCCAGATGTACACTAATGTGGACGAGGACCTCGTCGCTGGCAG 363
 QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
 DB 364 CGCGCCCGGGGGCGGTCTCTTGACACCATGACCTGTGGCAGCTCAGACCTTTACTTG 423
 QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
 DB 424 CTCACGAGACATGCTGACGTCATTCCGGTGGCGGGGGCGGACAGTAGGGGGAGCCTG 483
 QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerGlyGlyProLeuLeuCysPro 161
 DB 484 CTCCTCCCGCGGCGTCTCTCTACTTGAAGGGCTCTCGGGTGGTCCACTGCTCGCCT 543
 QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181
 DB 544 TCGGGGACGCTGTGGGCATCTCCGGCTGCGGCTGACCCCGGGGGTTGCAAGGCG 603
 QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196
 DB 604 GTGGACTTTGTGCGCGTAGATGCTCATGGAACACTACTATGCGGTCT 648

RESULT 14
 LOCUS AR145253 651 bp DNA linear PAT 08-AUG-2001
 DEFINITION Sequence 94 from patent US 6211338.
 ACCESSION AR145253
 VERSION AR145253.1 GI:15107120
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 651)
 AUTHORS Malcolm,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.
 TITLE Single-chain recombinant complexes of hepatitis C virus NS3

protease and NS4A cofactor peptide
Patent: US 6211338-A 94 03-APR-2001;

JOURNAL

FEATURES

source
Location/Qualifiers
1..651
/organism="unknown"
BASE COUNT 119 a 188 c 199 g 145 t
ORIGIN

Alignment Scores:

Pred. No.: 2,21e-66 Length: 651
Score: 900.50 Matches: 169
Percent Similarity: 94.36% Conservativeness: 15
Best Local Similarity: 86.67% Mismatches: 8
Query Match: 87.26% Indels: 3
DB: 6 Gaps: 1

US-09-965-594-16 (1-197) x AR145253 (1-651)

```
QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
|||||
Db 64 GGTTCGTGTTATTGTTGGTAGAATTATTTATCTGTTAGTGTATCATCGGCTAC 123
|||||
QY 22 AlaGlnGlnThrArgGlyGluGlyCysGlnGlnThrSerGlnThrGlyArgAspLys 41
|||||
Db 124 TCCCAACAGACGGCGGCTACTTGGTTGCATCAAGACTACAGCTTACAGCGCGGACAA 183
|||||
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
|||||
Db 184 AACCAAGTCGAGGAGAGTTTCAGGTGTTTCCACCGCAACACATCTTCTGGCGACC 243
|||||
QY 62 CysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
|||||
Db 244 TCGCTCAACGGCGTGTGTGGACGTTTACCATGGTGTGCTCAAGACCTTAGCGCGC 303
|||||
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
|||||
Db 304 CCAAGGGGCAATCACCAGATGATACATTAATGGACAGGACCTCGTGGCTGGCAG 363
|||||
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
|||||
Db 364 GCGCCCCCGGGCGGCTCTCTTGACACCATGCACCTGTGGCAGCTCAGACCTTTACTTG 423
|||||
QY 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141
|||||
Db 424 GTCACGACATGCTGACGCTATTCGGTGGCGCGGCGGCGAGACAGTAGGGGAGCCTG 483
|||||
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
|||||
Db 484 CTCTCCCCCAGGCGCTCTCTTACTTGAAGGGCTCTTCGGGTGGTCCACTGCTCTGCCCT 543
|||||
QY 162 AlaGlyHisAlaValGlyPheArgAlaAlaValCysThrArgGlyValAlaLysAla 181
|||||
Db 544 TCGGGGACGGCTGGGCACTCTCCGGCTGCGGCTGCGATGACCCCGGGGGTTCGGAAGCG 603
|||||
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196
|||||
Db 604 GTGGACTTGTGCGCGTAGAGTCCATCGAAACTACTATGCGGTCT 648
|||||
```

RESULT 15

AR145265
LOCUS 1998 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 106 from patent US 6211338.
ACCESSION AR145265
VERSION AR145265.1 GI:15107132

KEYWORDS

Unknown.
Organism.
Unclassified.

REFERENCE 1 (bases 1 to 1998)
AUTHORS Malcolm B.A., Taremi, S. Shane., Weber, P.C. and Yao, N.
TITLE Single-chain recombinant complexes of hepatitis C virus NS3
protease and NS4A cofactor peptide
JOURNAL Patent: US 6211338-A 106 03-APR-2001;

FEATURES
source
Location/Qualifiers
1..1998
/organism="unknown"

BASE COUNT 409 a 597 c 567 g 425 t
ORIGIN

Alignment Scores:

Pred. No.: 7,61e-66 Length: 1998
Score: 900.50 Matches: 169
Percent Similarity: 93.88% Conservativeness: 15
Best Local Similarity: 86.22% Mismatches: 9
Query Match: 87.26% Indels: 3
DB: 6 Gaps: 1

US-09-965-594-16 (1-197) x AR145265 (1-1998)

```
QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
|||||
Db 64 GGTTCGTGTTATTGTTGGTAGAATTATTTATCTGTTAGTGTATCATCGGCTAC 123
|||||
QY 22 AlaGlnGlnThrArgGlyGluGlyCysGlnGlnThrSerGlnThrGlyArgAspLys 41
|||||
Db 124 TCCCAACAGACGGCGGCTACTTGGTTGCATCATCTAGCTTACAGCGCGGACAA 183
|||||
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
|||||
Db 184 AACCAAGTCGAGGAGAGTTTCAGGTGTTTCCACCGCAACACATCTTCTGGCGACC 243
|||||
QY 62 CysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
|||||
Db 244 TCGCTCAACGGCGTGTGTGGACGTTTACCATGGTGTGCTCAAGACCTTAGCGCGC 303
|||||
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
|||||
Db 304 CCAAGGGGCAATCACCAGATGATACATTAATGGACAGGACCTCGTGGCTGGCAG 363
|||||
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
|||||
Db 364 GCGCCCCCGGGCGGCTCTCTTGACACCATGCACCTGTGGCAGCTCAGACCTTTACTTG 423
|||||
QY 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141
|||||
Db 424 GTCACGACATGCTGACGCTATTCGGTGGCGCGGCGGCGAGACAGTAGGGGAGCCTG 483
|||||
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
|||||
Db 484 CTCTCCCCCAGGCGCTCTCTTACTTGAAGGGCTCTTCGGGTGGTCCACTGCTCTGCCCT 543
|||||
QY 162 AlaGlyHisAlaValGlyPheArgAlaAlaValCysThrArgGlyValAlaLysAla 181
|||||
Db 544 TCGGGGACGGCTGGGCACTCTCCGGCTGCGGCTGCGATGACCCCGGGGGTTCGGAAGCG 603
|||||
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
|||||
Db 604 GTGGACTTGTGCGCGTAGAGTCCATCGAAACTACTATGCGGTCTCCG 651
|||||
```

Search completed: August 31, 2003, 00:46:14
Job time : 2570.57 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 30, 2003, 19:13:57 ; Search time 182.939 seconds
(without alignments)
2906.924 Million cell updates/sec

Title: US-09-965-594-16

Perfect score: 1032

Sequence: 1 MKKGSVVIVGRINLSGDTA.....VAKAVDFIPVESLETTMRSP 197

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-O=/cgn2.1/USPTO.spool/US09965594/runat_29082003.151918.28302/app_query.fasta_1.2872
-DB=N_Geneseq_15Jun03 -OFMT=fastap -SUFFIX=mg -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cd1
-LIST=45 -LOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09965594 -CGCN_1_1.1412 -runat_29082003.151918.28302 -NCPU=6 -ICPU=3
-NO_MMAPP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_19Jun03.*

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
25: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1032	100.0	594	21	AA73331	Hepatitis C virus
2	1015	98.4	594	21	AA73330	Hepatitis C virus
3	1002	97.1	594	21	AA73332	Hepatitis C virus
4	990	95.9	594	21	AA73333	Hepatitis C virus
5	981	95.1	588	21	AA73329	Hepatitis C virus
6	980	95.0	594	21	AA73334	Hepatitis C virus
7	976	94.6	594	21	AA73335	Hepatitis C virus
8	942	91.3	588	21	AA73328	Hepatitis C virus
9	922.5	89.4	12734	24	ABA95615	Chimeric BVDV/HCV
10	911.5	88.3	1998	20	AA80355	HCV NS4A-NS3 compl
11	908.5	88.0	1998	20	AA80359	HCV NS4A-NS3 compl
12	907.5	87.9	1998	20	AA80353	HCV NS4A-NS3 compl
13	907.5	87.9	1998	20	AA80354	HCV NS4A-NS3 compl
14	904.5	87.6	612	25	ABX15706	Anti-viral synthet
15	904.5	87.6	651	20	AA80345	HCV NS4A-NS3 compl
16	904.5	87.6	1998	20	AA80357	HCV NS4A-NS3 compl
17	904.5	87.6	1998	20	AA80358	HCV NS4A-NS3 compl
18	903.5	87.5	1998	20	AA80352	HCV NS4A-NS3 compl
19	903.5	87.5	2013	20	AA80360	HCV NS4A-NS3 compl
20	901.5	87.4	651	20	AA80349	HCV NS4A-NS3 compl
21	900.5	87.3	651	20	AA80343	HCV NS4A-NS3 compl
22	900.5	87.3	651	20	AA80344	HCV NS4A-NS3 compl
23	900.5	87.3	1998	20	AA80356	HCV NS4A-NS3 compl
24	900.5	87.3	2016	20	AA80361	HCV NS4A-NS3 compl
25	900	87.2	648	20	AA80365	HCV NS4A-NS3 compl
26	898	87.0	648	20	AA80363	HCV NS4A-NS3 compl
27	897.5	87.0	650	20	AA80347	HCV NS4A-NS3 compl
28	897.5	87.0	651	20	AA80348	HCV NS4A-NS3 compl
29	897.5	87.0	651	20	AA80351	HCV NS4A-NS3 compl
30	896.5	86.9	651	20	AA80342	HCV NS4A-NS3 compl
31	894	86.6	648	20	AA80362	HCV NS4A-NS3 compl
32	893.5	86.6	650	20	AA80346	HCV NS4A-NS3 compl
33	893.5	86.6	651	20	AA80350	HCV NS4A-NS3 compl
34	891	86.3	8145	20	AA823259	Plasmid pET-BS(+)/HCV NS3 DNA. Hepa
35	889	86.1	1933	20	AA823258	Hepatitis C virus
36	888.5	86.1	9646	19	AAV59361	cDNA encoding hepa
37	888.5	86.1	9646	24	ABR87285	Hepatitis C virus
38	888.5	86.1	12980	19	AAV59364	Hepatitis C virus
39	888.5	86.1	12980	24	ABR87286	Hepatitis C virus
40	888.5	86.1	16622	21	AA236212	Nucleotide sequenc
41	884.5	85.7	5300	10	AA92097	Combined open read
42	884.5	85.7	5360	10	AA92027	Hepatitis C virus
43	884.5	85.7	6905	10	AA92103	Combined open read
44	884.5	85.7	7310	10	AA92106	Combined open read
45	884.5	85.7	7310	10	AA92036	Composite hepatiti

ALIGNMENTS

RESULT 1
AA73331
ID AAA73331 standard; DNA; 594 BP.
XX
XX
AC AAA73331;
XX
DT 19-DEC-2000 (first entry)
XX
DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #4.
XX
XX Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW liver failure; liver cancer; mutant; mutain; ds.
XX
XX Hepatitis C virus.
OS Synthetic.
XX
XX
FH Key Location/Qualifiers

```

CDS      1..594
        /*tag= a
        /product= "NS4A-NS3 fusion protein #4"

WO200040707-A1.
13-JUL-2000.
06-JAN-2000; 2000WO-US00345.
08-JAN-1999; 99US-0115271.
(BRIM ) BRISTOL-MYERS SQUIBB CO.
Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
WPI: 2000-465976/40.
P-PSTDB; AAB15222.

Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
amino acid, useful for screening inhibitors that may treat hepatitis C
X_
XX
Claim 26; Fig 14; 66pp; English.

The present sequence is the coding sequence for a mutated version of a
fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
protease enzymes. These proteins are both essential for the replication
of the virus, acting to cleave its replicative proteins from the
CC polyprotein produced from the HCV genome. Inhibitors of the two proteins
should be effective as antiviral treatments of HCV infection. This is
useful as HCV can lead to chronic liver disease such as cirrhosis, liver
failure and liver cancer. The present invention concerns a number of NS3
mutants and NS3-NS4A fusion proteins which can be used to identify
inhibitors of this type, as well as enabling structural studies of the
protease and protease-inhibitor complexes. The protein produced from this
sequence contains the alpha-helix0-1 variant.
XX
SQ Sequence 594 BP; 105 A; 187 C; 155 G; 147 T; 0 other;

Alignment Scores:
Pred. No.:            8,48e-88          Length:           594
Score:                1032.00         Matches:           197
Percent Similarity:   100.00%       Conservative:    0
Best Local Similarity: 100.00%     Mismatches:      0
Query Match:         100.00%       Indels:          0
DB:                  21             Gaps:            0

US-09-965-594-16 (1-197) x AAA73331 (1-594)

Qy      1 MetLysLYSLysGlySerValIleValGlyArgTlleAsnLeuSerGlyAspThrAla 20
Db      1 ATCAAAAAAAAAGGATCGCGTTGTATTCTGTCGGCGTGATCAAACCTGTCCGGTGACACGGT 60
Qy     21 TyrAlaglnGlnThrArgGlycyluciluglyCySGlngluThrSerGlnThrGlyArGasg 40
Db     61 TAGCCTCACACACTCGAGGTGAGGAGGTGCCAAAGAACCCTCCCAGACCGGTGCGTAGC 120
Qy     41 LysAsnClnVaiglcglucluglvalGlnilleValSerThralatThrGlnInthrPhelAuLa 60
Db    121 ANAACCAAGTTGAAGTTGAAGTTGCAGATCGTTCCACCCTACCCAGACCTTCCCTGGCT 180
Qy     61 ThrCysIIeaSncglyvaICystTrpthrvValTyrtHisGIylaGlYthrArGThrIleAla 80
Db    181 ACCTGCTATCAAGCGTGTITTCGTGGACCGGTATTACACACGGTGCTGTGTACCCGTACCATCGCT 240
Qy     81 SerProLyScglYProvallThrGlnMetFyrThrasNValasPLysAspLeuVaiGLyTip 100
Db    241 TCCECGAAGAGGTCGCGTTACCCAGATGTACACCAACGTTGACAAGACCTTGGTTGGTTGG 100
Qy    101 GluAlaPrOgIncnglySerArgserLeuthrProcYsthyrCysGlyserySerAspleutyir 120

```

xx SQ Sequence 594 BP; 103 A; 186 C; 156 G; 149 T; 0 other;

Alignment Scores:

Pred. No.:	3,33e-86	Length:	594
Score:	1015.00	Matches:	194
Percent Similarity:	98.48%	Conservative:	0
Best Local Similarity:	98.48%	Mismatches:	3
Query Match:	98.35%	Indels:	0
DB:	21	Gaps:	0

US-09-965-594-16 (1-197) x AAA73332 (1-594)

QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
 DB 1 ATGAATAAAAGGATCGTGTATCGTCGCCGTATCAACCTGTCGGTGACACCGCT 60
 QY 21 TyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAsp 40
 DB 61 TACGCTCAGCAGACTCGAGGTCAGGAGGTTGCCAAGAAACCTCCACAGCCGTCGTGAC 120
 QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
 DB 121 AAAAACAGGTTGAAGTCAAGTTCAGATCGTTTCCACCGCTGCTCAGACCTTCTCGCT 180
 QY 61 ThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrIleAla 80
 DB 181 ACCTGCATCAACGGTGTGTGTCGACCGTTTACCACGGTGTGTGTCGCTACCATCGCT 240
 QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr 100
 DB 241 TCCCGGAAAGTCCGGTTATCCAGATGACACCAACGTTGACAAAGACCTCGTTGGTGG 300
 QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
 DB 301 CCGGCTCCGAGGGTTCGGTTCCTGACCCCGTGCACCTCGGTTCTCCGACCTGTAC 360
 QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
 DB 361 CTGGTTACCGCTCAGCTGACGTTATCCCGTTCGTCGTCGTCGTCGTCGTCGTCGTCG 420
 QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
 DB 421 CTGCTGTCCCGCGTCCGATCTCTTACCTGAAGGTTCCCTCCGGTGGTCCGCTGTGTC 480
 QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys 180
 DB 481 CCGGCTGCTCAGCTGTGTGTATCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
 QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
 DB 541 GCTGTTGACTTCATCCCGTGAATCCCTGGAAACACCATCGCTGCCCG 591

RESULT 3

ID AAA73332 standard; DNA; 594 BP.

XX AC AAA73332;

XX DT 19-DEC-2000 (first entry)

DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #5.

XX Hepatitis; NS3 protease; viral replication; chronic liver disease;
 KW liver failure; liver cancer; mutant; mutain; ds.

XX Hepatitis C virus.

OS Synthetic.

XX Key Location/Qualifiers
 FH 1..594
 CDS /tag= a

FT /product= "NS4A-NS3 fusion protein #5"

xx WO200040707-A1.
 xx 13-JUL-2000.
 xx 06-JAN-2000; 2000WO-0500345.
 xx 08-JAN-1999; 990S-0115271.
 xx (BRIM) BRISTOL-MYERS SQUIBB CO.
 xx Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
 PI WPI: 2000-465976/40.
 DR P-PSDB; AAB15223.
 xx Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 PT .
 xx Claim 26; Fig 15; 66pp; English.
 xx The present sequence is the coding sequence for a mutated version of a
 CC fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
 CC protease enzymes. These proteins are both essential for the replication
 CC of the virus, acting to cleave its replicative proteins from the
 CC polyprotein produced from the HCV genome. Inhibitors of the two proteins
 CC should be effective as antiviral treatments of HCV infection. This is
 CC useful as HCV can lead to chronic liver disease such as cirrhosis, liver
 CC failure and liver cancer. The present invention concerns a number of NS3
 CC mutants and NS3-NS4A fusion proteins which can be used to identify
 CC inhibitors of this type, as well as enabling structural studies of the
 CC protease and protease-inhibitor complexes. The protein produced from this
 CC sequence contains the alpha-helix0-1 variant.
 xx Sequence 594 BP; 105 A; 189 C; 153 G; 147 T; 0 other;

Alignment Scores:

Pred. No.:	5.51e-85	Length:	594
Score:	1002.00	Matches:	194
Percent Similarity:	98.48%	Conservative:	0
Best Local Similarity:	98.48%	Mismatches:	3
Query Match:	97.09%	Indels:	0
DB:	21	Gaps:	0

US-09-965-594-16 (1-197) x AAA73332 (1-594)

QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
 DB 1 ATGAATAAAAGGATCGTGTATCGTCGCCGTATCAACCTGTCGGTGACACCGCT 60
 QY 21 TyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAsp 40
 DB 61 TACGCTCAGCAGACTCGAGTGAAGTTCAGATCGTTTCCACCGCTTCCACAGACCTTCTCGCT 120
 QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
 DB 121 AAAAACAGGTTGAAGTGAAGTTCAGATCGTTTCCACCGCTTCCACAGACCTTCTCGCT 180
 QY 61 ThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrIleAla 80
 DB 181 ACCTCCATCAACGGTGTGTGTCGACCGTTTACCACGGTGTGTGTCGCTACCATCGCT 240
 QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr 100
 DB 241 TCCCGGAAAGTCCGGTTACCCAGATGACACCAACGTTCCAAAGACCTTGGTGGTGG 300
 QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
 DB 301 CAGGCTCCGAGGGTTCGGTTCCTGACCCCGTGCACCTCGCGTTCCTCCGACCTGTAC 360
 QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140

PD 13-JUL-2000.
 XX
 PF 06-JAN-2000; 2000WO-US00345.
 XX
 PR 08-JAN-1999; 99US-0115271.
 XX
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX
 PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
 XX
 DR WPI: 2000-465976/40.
 DR P-PSDB; AAB15220.
 DR
 XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 PT
 XX
 XX Claim 26; Fig 12; 66pp; English.
 PS
 XX The present sequence is the coding sequence for a mutated version of a
 CC fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
 CC protease enzymes. These proteins are both essential for the replication
 CC of the virus, acting to cleave its replicative proteins from the
 CC polyprotein produced from the HCV genome. Inhibitors of the two proteins
 CC should be effective as antiviral treatments of HCV infection. This is
 CC useful as HCV can lead to chronic liver disease such as cirrhosis, liver
 CC failure and liver cancer. The present invention concerns a number of NS3
 CC mutants and NS3-NS4A fusion proteins which can be used to identify
 CC inhibitors of this type, as well as enabling structural studies of the
 CC protease and protease-inhibitor complexes. The protein produced from this
 CC sequence contains the alpha-helix0-1 variant.
 XX
 XX Sequence 588 BP; 103 A; 180 C; 156 G; 149 T; 0 other;
 SQ

Alignment Scores:
 Pred. No.: 5,06e-83 Length: 588
 Score: 981.00 Matches: 190
 Percent Similarity: 96.95% Conservative: 1
 Best Local Similarity: 96.45% Mismatches: 4
 Query Match: 95.06% Indels: 2
 DB: 21 Gaps: 1

US-09-965-594-16 (1-197) x AAA73329 (1-588)

2y 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
 Db 1 ATGAAATAAAAGATCCGTTGTTATCTGCGGCCGATATAGTACTGACCGT-----GCT 54
 Qy 21 TyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAsp 40
 Db 55 TACGCTCAGCAGACTCGAGGTGAGGAGGTTGCCAAGAAACCTCCACAGACCGCTCGTAC 114
 Qy 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
 Db 115 AAAAACAGGTTGAAGGTGAAGTTCAGATCGTTCCACCGCTGCTCAGACCTTCCTGGCT 174
 Qy 61 ThrCysIleAsnGlyValCysTrpThrValThrHisGlyAlaGlyThrArgThrIleAla 80
 Db 175 ACCTGCATCAACGGTGTTCGTCGACCGTTTACACCGGTCTGCTGCTACCGGTACCTCGCT 234
 Qy 81 SerProLysGlyProValThrGlnMetThrThrAsnValAspLysAspLeuValGlyTrp 100
 Db 235 TCCCGAAAGGTCCGGTTATCCAGATGATACCAACGTTGACAAAGACCTGGTGGTGGTGG 294
 Qy 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuThr 120
 Db 295 CCGGCTCGCAGAGGTTCCCGTTCCTGACCCCGGTGACCTGCGGTTCCTCCGACCTGTAC 354
 Qy 121 LeuValThrArgHisAlaValIleProValArgArgGlyAspSerArgGlySer 140
 Db 355 CTGGTTACCCGTCACGCTGAGCTTATCCCGGTTTCGTCGTCGTCGTCGTCGTCGTCGTC 414

Qy 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
 Db 415 CTGCTGTCCCGCGTCCGATCTCCACCTGAAGGTTCCTCCGGTGGTCCGCTGCTGTGC 474
 Qy 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys 180
 Db 475 CCGGCTGGTCACGCTGTGTGATCTTCGCTGCTGCTGTTGACCCCGTGGTGTCTAA 534
 Qy 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
 Db 535 GCTGTTGACTTCATCCCGGTTGAATCCCTGGAAACACCATCGCTTCCCGC 585

RESULT 6
 AAA73334
 ID AAA73334 standard; DNA; 594 BP.
 XX
 AC AAA73334;
 XX
 DT 19-DEC-2000 (first entry)
 XX
 DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #7.
 XX
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
 KW liver failure; liver cancer; mutant; mutain; ds.
 XX
 OS Hepatitis C virus.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..594
 FT /*tag= a
 FT /product= "NS4A-NS3 fusion protein #7"
 XX
 PN WO200040707-A1.
 XX
 PD 13-JUL-2000.
 XX
 PF 06-JAN-2000; 2000WO-US00345.
 PR 08-JAN-1999; 99US-0115271.
 XX
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX
 PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
 XX
 DR WPI: 2000-465976/40.
 DR P-PSDB; AAB15225.
 XX
 PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 PT
 XX
 XX Claim 26; Fig 17; 66pp; English.
 PS
 XX The present sequence is the coding sequence for a mutated version of a
 CC fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
 CC protease enzymes. These proteins are both essential for the replication
 CC of the virus, acting to cleave its replicative proteins from the
 CC polyprotein produced from the HCV genome. Inhibitors of the two proteins
 CC should be effective as antiviral treatments of HCV infection. This is
 CC useful as HCV can lead to chronic liver disease such as cirrhosis, liver
 CC failure and liver cancer. The present invention concerns a number of NS3
 CC mutants and NS3-NS4A fusion proteins which can be used to identify
 CC inhibitors of this type, as well as enabling structural studies of the
 CC protease and protease-inhibitor complexes. The protein produced from this
 CC sequence contains the alpha-helix0-7 variant.
 XX
 XX Sequence 594 BP; 105 A; 192 C; 151 G; 146 T; 0 other;
 SQ

Alignment Scores:
 Pred. No.: 6.36e-83 Length: 594
 Score: 980.00 Matches: 190

Percent Similarity: 97.46% Conservative: 2
 Best Local Similarity: 96.45% Mismatches: 5
 Query Match: 94.96% Indels: 0
 DB: 21 Gaps: 0

US-09-965-594-16 (1-197) x AAA73334 (1-594)

QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
 DB 1 ATGAAAAAAGAGATCCGTTGTTATCGTGGCCGTATCAACCTGTCGGTGACACCGCT 60
 QY 21 TyrAlaGlnGlnThrArgGlyGluGlyCysGlnGlnThrSerGlnThrGlyArgasp 40
 DB 61 TAGCGTCAGCAGACTCGAGGTCAGCAGGATACCCAGAGACCTCCACACCGGTGCTGAC 120
 QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
 DB 121 AAAACCCAGGTTGAAGTGAAGTTCAGATCGTTTCCACCGCTACCCAGCTTCTGGCT 180
 QY 61 ThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
 DB 181 ACCTCCATCAACGGTGTCTGTGGACCGTTTACCACGGTGTGGTACCGTACCATCGCT 240
 QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp 100
 DB 241 TCCCGGAAAGGTCGGGTTACCCAGATGTACACCAACGTTGACAAAGACCTGGTGGTGG 300
 QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
 DB 301 CAGGCTCCGACAGGTTCCGCTCCCGTCCCGCTGACCTGGGTTCTCCGACCTGTAC 360
 QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
 DB 361 CTGGTTACCGCTCACGCTGACGTTATCCCGGTTCTGCTGCTGCTGCTGCTGCTGCT 420
 QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
 DB 421 CTGCTGTCCCGGTCGGATCTCTACCTGAAGGTTCTCCCGTGGTCCGCTGCTGTC 480
 QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys 180
 DB 481 CCGGCTGGTCACGCTGTGTGTATCTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
 QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
 DB 541 GCTGTTGACTTATCCCGGTTGAATCCCTGGAACACCATCGTTCGCCG 591

RESULT 7

AAA73335
 ID AAA73335 standard; DNA: 594 BP.

AC AAA73335;

XX 19-DEC-2000 (first entry)

XX Hepatitis C virus NS4A-NS3 fusion protease coding sequence #8.

XX Hepatitis; NS3 protease; viral replication; chronic liver disease;

XX liver failure; liver cancer; mutant; mutein; ds.

XX Hepatitis C virus.

OS Synthetic.

XX Key Location/Qualifiers

FT CDS

FT 1..594

FT /*tag- a

FT /product- "NS4A-NS3 fusion protein #8"

XX W0200040707-A1.

XX 13-JUL-2000.

XX 06-JAN-2000; 2000WO-US000345.

XX 08-JAN-1999; 99US-0115271.
 XX (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;

XX WPI: 2000-465976/40.
 XX P-PSDB; AAB15226.

XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 XX substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 XX amino acid, useful for screening inhibitors that may treat hepatitis C

XX Disclosure; Fig 18; 66pp; English.

XX The present sequence is the coding sequence for a mutated version of a
 CC fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
 CC protease enzymes. These proteins are both essential for the replication
 CC of the virus, acting to cleave its replicative proteins from the
 CC polypeptide produced from the HCV genome. Inhibitors of the two proteins
 CC should be effective as antiviral treatments of HCV infection. This is
 CC useful as HCV can lead to chronic liver disease such as cirrhosis, liver
 CC failure and liver cancer. The present invention concerns a number of NS3
 CC mutants and NS3-NS4A fusion proteins which can be used to identify
 CC inhibitors of this type, as well as enabling structural studies of the
 CC protease and protease-inhibitor complexes. The protein produced from this
 CC sequence contains the alpha-helix0 wild-type sequence.

XX Sequence 594 BP; 98 A; 189 C; 153 G; 154 T; 0 other;

XX Alignment Scores:

XX Pred. No.: 1.51e-82 Length: 594
 XX Score: 976.00 Matches: 189
 XX Percent Similarity: 95.94% Conservative: 0
 XX Best Local Similarity: 95.94% Mismatches: 8
 XX Query Match: 94.57% Indels: 0
 XX DB: 21 Gaps: 0

XX US-09-965-594-16 (1-197) x AAA73335 (1-594)

QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
 DB 1 ATGAAAAAAGAGATCCGTTGTTATCGTGGCCGTATCAACCTGTCGGTGACACCGCT 60
 QY 21 TyrAlaGlnGlnThrArgGlyGluGlyCysGlnGlnThrSerGlnThrGlyArgasp 40
 DB 61 TAGCGTCAGCAGACTCGAGGTCGCTGGTGTGATCATCATCCCTGACCGCTGCTGAC 120
 QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
 DB 121 AAAACCCAGGTTGAAGTGAAGTTCAGATCGTTTCCACCGCTGCTCAGACCTTCTGGCT 180
 QY 61 ThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
 DB 181 ACCTCCATCAACGGTGTCTGTGGACCGTTTACCACGGTGTGGTACCGTACCATCGCT 240
 QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp 100
 DB 241 TCCCGGAAAGGTCGGGTTATCCAGATGTACACCAACGTTGACAAAGACCTGGTGGTGG 300
 QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
 DB 301 CAGGCTCCGACAGGTTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGT 360
 QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
 DB 361 CTGGTTACCGCTCACGCTGACGTTATCCCGGTTCTGCTGCTGCTGCTGCTGCTGCT 420
 QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
 DB 421 CTGCTGTCCCGGTCGGATCTCTACCTGAAGGTTCTCCCGTGGTCCGCTGCTGTC 480

QY 161 ProAlaGlyHisAlaValAlcIlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys 180
 DB 481 CCGGTGTCACGCGTGGGTATCTTCGCTGCTGCTTTTGCACCGCTGTGTGCTAA 540
 QY 181 AlaValAspPheIleProValcIlySerLeuGluThrThrMetArgSerPro 197
 DB 541 GCTGTTGACTTCATCCCGTGTGAATCCCTGGAAACCAACCATGCGTTCCTCCG 591

RESULT 8 AAA73328

ID AAA73328 standard; DNA; 588 BP.

AC AAA73328;

DT 19-DEC-2000 (first entry)

DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #1.

KW Hepatitis; NS3 protease; viral replication; chronic liver disease;

KW liver failure; liver cancer; ds.

XX Hepatitis C virus.

OS Synthetic.

Key Location/Qualifiers

FT 1..588

FT /tag= a

FT /product= "NS3-NS4A fusion protein"

XX WO200040707-A1.

PN 13-JUL-2000.

XX 06-JAN-2000; 2000WO-US00345.

XX 08-JAN-1999; 99US-0115271.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

PA Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;

XX WPI; 2000-465976/40.

XX P-PSDB; AAB15212.

XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 PT -

PS Disclosure; Fig 10; 66pp; English.

XX The present sequence is the coding sequence for a fusion protein created
 CC using the hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
 CC proteins are both essential for the replication of the virus, acting to
 CC cleave its replicative proteins from the polyprotein produced from the
 CC HCV genome. Inhibitors of the two proteins should be effective as
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A
 CC fusion proteins which can be used to identify inhibitors of this type, as
 CC well as enabling structural studies of the protease and
 CC protease:inhibitor complexes.

XX Sequence 588 BP; 97 A; 183 C; 153 G; 155 T; 0 other;

Alignment Scores:

Pred. No.:	2,29e-79	Length:	588
Score:	942.00	Matches:	185
Percent Similarity:	94.42%	Conservative:	1
Best Local Similarity:	93.91%	Mismatches:	2
Query Match:	91.28%	Indels:	2
DB:	21	Gaps:	1

US-09-965-594-16 (1-197) x AAA73328 (1-588)

QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
 DB 1 ATGAAAAAAGAGTTCCGTTGTTATCGTCGCCCGCTATAGTACTGAACGGT-----GCT 54
 QY 21 TTTATAGInGInThrArgGlyGluGluGlyCysGInGluThrSerGInThrGlyArgASP 40
 DB 55 TAGCTCAGCAGACTCGAGGTCGTGGTTGCATCATCACCCTCCCTGACCGGTGCTGAC 114
 QY 41 LysAsnGInValGluGluValGInIleValSerThrAlaThrGInThrPheLeuAla 60
 DB 115 AAAACACAGGTTGAAGGTGAAGTTCAGATCGTTCCACCGCTGCTCAGACCTTCTCGCT 174
 QY 61 ThrCysIleAsnGlyValCysTrpThrValThrHisGlyAlaGlyThrArgThrIleAla 80
 DB 175 ACCTGCATCAACGGTGTTCGTGGACCGTTTACCACGGTGTGTACCGGTACCATCGCT 234
 QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValcIlyTrp 100
 DB 235 TCCCGAAGGTCGCGTTATCCAGATGTACACCAAGCTTGACAAAGACCTGTTGGTTGG 294
 QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
 DB 295 CCGGCTCCGACAGGTTCCGTTCCCTGACCCCGTCACCTCGGTTCTCCGACCTGTAC 354
 QY 121 LeuValThrArgHisAlaAspValIleProValAlaArgArgGlyAspSerArgGlySer 140
 DB 355 CTGGTTACCCGTCACGCTGACGTTATCCCGGTTCTCGTGGTGACCTCCGTTGGTTC 414
 QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
 DB 415 CTGCTGTCGCCGCGTCGATCTCTACCTGAAAGGTTCTCCGTTGGTCCGCTGCTGTC 474
 QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys 180
 DB 475 CCGGCTGGTCACGCTGTTGGTATCTCCGTTGCTGCTGTTGCACCGGTGTTGCTAAA 534
 QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
 DB 535 GCTGTTGACTTCATCCCGTGTGAATCCCTGGAACCAACCATGCGTTCCTCCG 585

RESULT 9

ABA95615

ID ABA95615 standard; DNA; 12734 BP.

XX AC ABA95615;

XX AC ABA95615;

DT 21-MAR-2002 (first entry)

DE Chimeric BVDV/HCV NS3-wt sequence.

XX Pestivirus; Npro; protease; NS3; screening; ds.

XX Chimeric - Bovine viral diarrhea virus.

OS Chimeric - Hepatitis C virus.

XX US6326137-B1.

XX 04-DEC-2001.

XX 25-JUN-1999; 99US-0344456.

XX 25-JUN-1999; 99US-0344456.

XX (SCHE) SCHERING CORP.

XX Hong Z, Lai VCH, Lau JYN;

XX WPI; 2002-121103/16.

PT Nucleic acid construct encoding chimeric hepatitis C virus (HCV)

PT pestivirus genome where the Npro protease gene is replaced with NS3
 PT protease gene, useful for in vivo screening of compounds which inhibit
 XX HCV infection

XX Example 2: Columns 17-28; 20pp; English.

CC The present invention relates to a nucleic acid construct encoding a
 CC chimeric Hepatitis C virus (HCV)-pestivirus genome. The construct
 CC comprises a pestivirus genome where a Npro pestivirus protease gene is
 CC replaced with a gene encoding a functional HCV NS3 protease. Furthermore,
 CC each junction site recognised by the Npro protease is replaced with a
 CC junction site recognised by the HCV NS3 protease. The construct is useful
 CC for screening compounds that inhibit HCV in vivo by inhibiting HCV
 CC protease, where screening may be in cell culture or in an animal model.
 CC The present sequence is a chimeric clone of BVDV (bovine diarrhoea
 CC virus)/HCV NS3-wt, which was used to illustrate the present invention.

XX SQ Sequence 12734 BP; 4032 A; 2604 C; 3295 G; 2803 T; 0 other;

Alignment Scores:
 Pred. No.: 7.1e-76 Length: 12734
 Score: 922.50 Matches: 180
 Percent Similarity: 94.36% Conservativity: 4
 Best Local Similarity: 92.31% Mismatches: 8
 Query Match: 89.39% Indels: 3
 DB: 24 Gaps: 1

US-09-965-594-16 (1-197) x ABA95615 (1-12734)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
 DB 413 GGTAGTGTGTATTTGTTGTTAGTATGTTTATCTGTTAGTGTAGTATCATCGCGGTAC 472
 QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGlnThrSerGlnThrGlyArgAspLys 41
 DB 473 GCCCAGCAGCAGCAGCAGCCTCTAGGTGTAGATCACCAGTCTGACTGCGCGGGACAAA 532
 QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
 DB 533 AACCAAGTGGAGGTGGTGGTCCAGATGCTCAACTGCTACCCCAACCTTCTCTGGCAAG 592
 QY 62 CysIleAsnGlyValCysTyrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
 DB 593 TGCATCAATGGGTATGCTGGAGTCTCTACACCGGGCCGACGAGGACCATCGCATCA 652
 QY 82 ProGlyGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101
 DB 653 CCCAAGGTCTCTGTCATCCAGATGTATACCAATGTGGACCAAGACCTTGTGGCTGGCC 712
 QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
 DB 713 GCTCTCAAGGTTCCTGCTCATTTGACACCTGACCTGCGGCTCTCTCGGACCTTTACCTG 772
 QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
 DB 773 GTTAGGAGCAGCGCGAGCTCATTCCTGCTGCGCGGAGGTATAGCAGGGGTAGCCTG 832
 QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
 DB 833 CTTTGGCCCGCCGCTTCTACTATAAGGCTCTCTCGGGGGGCTCGCTGTGTGCCCC 892
 QY 162 AlaGlyHisAlaValAlaGlyIlePheAlaAlaValCysThrArgGlyValAlaLysAla 181
 DB 893 CGGGGACACCGCGTGGGCTATTTCAGGGCCGCGTGTGCACCCGTGGAGTGGCAAGCG 952
 QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196
 DB 953 GTGGACTTATCCCTGTGGAGAAGCTAGAGACAACCATGATGATCC 997

RESULT 10

AXX80355

ID AAX80355 standard; cDNA; 1998 BP.

XX

AC AAX80355;
 XX 07-SEP-1999 (first entry)
 DT
 XX HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:105.
 DE
 XX HCV; hepatitis C virus; single chain recombinant complex; linker;
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
 KW hydrophobic domain; covalent complex; detection; inhibitor; ss.
 XX Hepatitis C virus.
 OS Synthetic.
 XX WO9928482-A2.
 PN 10-JUN-1999.
 PD
 XX 24-NOV-1998; 98WO-US24528.
 PF
 XX 28-JUL-1998; 98US-0094331.
 PR 28-NOV-1997; 97US-0067315.
 XX (SCHE) SCHERING CORP.
 PA Malcolm BA, Taremi SS, Weber PC, Yao N;
 PI WPI; 1999-385385/32.
 XX New hepatitis C virus covalent complexes
 PT
 XX Disclosure; Page 166-169; 21pp; English.
 PS
 XX The present invention describes a covalent hepatitis C virus (HCV)
 CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
 CC to the amino terminus of the HCV NS3 protease domain. The present
 CC sequence encodes an example of the above complex. The covalent
 CC NS4A-NS3 complexes are useful for structural determination and
 CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.
 CC They can also be used for detecting inhibitors of the protease activity,
 CC the helicase activity and the ATPase activity of NS3. The covalent
 CC NS4A-NS3 complexes are more soluble, stable and active than the non-
 CC covalent protease-peptide complexes previously available.
 XX SQ Sequence 1998 BP; 411 A; 595 C; 569 G; 423 T; 0 other;

Alignment Scores:
 Pred. No.: 7.6e-76 Length: 1998
 Score: 911.50 Matches: 170
 Percent Similarity: 94.90% Conservativity: 16
 Best Local Similarity: 86.73% Mismatches: 7
 Query Match: 88.32% Indels: 3
 DB: 20 Gaps: 1

US-09-965-594-16 (1-197) x AAX80355 (1-1998)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
 DB 64 GGTCTGTGTATTTGTTGTTAGTATTTATTTATCTGTTAGTGTAGTATCATCGGCTAC 123
 QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGlnThrSerGlnThrGlyArgAspLys 41
 DB 124 TCCCAACAGACGCGGGGCTACTTGTGTTGCAAGAGACTAGCCTTACAGCGCGGGACAG 183
 QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
 DB 184 ACCAGGTGCGAGGAGAGGTTCAGTGTGTTTCCACCGCAACAACTCTCTCTGGGACC 243
 QY 62 CysIleAsnGlyValCysTyrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
 DB 244 TCGGTCAACGCGGTGTGTGGACCTTTACCTGCTGGCTCAAGACCTTAGCGGC 303

QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
 DB 304 CCAAAAGGGCCCAATCACCAGATGACACTAATGTGGACCAGGACCTGCTGGCTGGCAG 363
 QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
 DB 364 GCGCCCGCCGGGGCGCTCTTGCACACCATGACCTGTGGCAGCTCAGACCTTACTTG 423
 QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
 DB 424 GTCACGAGACATGCTGAGCTCATTCGGGTGCGCGGGGGCGGACAGTAGGGGAGCCTG 483
 QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
 DB 484 CTCCTCCCGGAGCCCTGCTCTCTACTTGAAGGCTCTTCGGGTGTCCACCTCTGCCCC 543
 QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181
 DB 544 TCGGGGACGCTGTGGCATCTTCGGGCTGCCGTATGCACCGGGGGGTTCGGAAGCG 603
 QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
 DB 604 GTGGACTTTGTGCCGTAGAGTCCATGGAACTACTATGCGGTCTCCG 651

RESULT 11

AX80359
 ID AX80359 standard; cDNA; 1998 BP.

XX AX80359;
 AC AX80359;
 XX AX80359;
 DT 07-SEP-1999 (first entry)
 XX HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:109.

DE HCV; hepatitis C virus; single chain recombinant complex; linker;
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
 KW hydrophobic domain; covalent complex; detection; inhibitor; ss.

XX Hepatitis C virus.
 OS Synthetic.

XX W09928482-A2.

XX 10-JUN-1999.

XX 24-NOV-1998; 98WO-US24528.

XX 28-JUL-1998; 98US-0094331.

PR 28-NOV-1997; 97US-0067315.

XX (SCHE) SCHERING CORP.

XX Malcolm BA, Taremi SS, Weber PC, Yao N;

XX WPI; 1999-385385/32.

XX New hepatitis C virus covalent complexes

XX Disclosure; Page 179-182; 211pp; English.

CC The present invention describes a covalent hepatitis C virus (HCV)
 CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
 CC to the amino terminus of the HCV NS3 protease domain. The present
 CC sequence encodes an example of the above complex. The covalent
 CC NS4A-NS3 complexes are useful for structural determination and
 CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.
 CC They can also be used for detecting inhibitors of the protease activity,
 CC the helicase activity and the ATPase activity of NS3. The covalent
 CC NS4A-NS3 complexes are more soluble, stable and active than the non-
 CC covalent protease-peptide complexes previously available.

SQ Sequence 1998 BP; 411 A; 595 C; 569 G; 423 T; 0 other;

Alignment Scores:

Pred. No.: 1.45e-75 Length: 1998
 Score: 908.50 Matches: 169
 Percent Similarity: 94.90% Conservative: 17
 Best Local Similarity: 86.22% Mismatches: 7
 Query Match: 88.03% Indels: 3
 DB: 20 Gaps: 1

US-09-965-594-16 (1-197) x AX80359 (1-1998)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
 DB 64 GGTTCGTGTTGTTGTTGGTAGAATATTATTCTGGTAGTGTATACCGCCCTAC 123
 QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGlnThrSerGlnThrGlyArgAspLys 41
 DB 124 TCCCAACACACGCGGGGCTACTTGGTTGCAAGAAGACTAGCTTACAGGCGGGCAAG 183
 QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
 DB 184 AACCAAGTTCGAGGGAGAGTTCCAGGTGGTTTCCACCGCAACACATCTCTCTGGCGACC 243
 QY 62 CysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
 DB 244 TCGCTCAACGCGCTGTGTTGGACCGTTTACCATGCTGCTGCTCAAGACCTTAGCCGCGC 303
 QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
 DB 304 CCAAAAGGGCCCAATCACCAGATGACACTAATGTGGACCAAGACCTCTGCGGTGGCAG 363
 QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
 DB 364 GCGCCCGCCGGGGCGCTTCTTGACACCTGACCTGTGGCAGCTCAGACCTTACTTG 423
 QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
 DB 424 GTCACGAGACATGCTGACGTCTATTCGGTGCAGCGGGGGCGGACAGTAGGGGAGCCTG 483
 QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
 DB 484 CTCCTCCCGGAGCCCTGCTCTCTACTTGAAGGCTCTGCTGGTGGTCCACTGCTCTGCCCT 543
 QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181
 DB 544 TCGGGGACGCTGTGGCATCTTCGGGCTGCCGTATGCACCGGGGGGTTCGGAAGCG 603
 QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
 DB 604 GTGGACTTTGTGCCGTAGAGTCCATGGAACTACTATGCGGTCTCCG 651

RESULT 12

AX80353

ID AX80353 standard; cDNA; 1998 BP.

XX AX80353;
 AC AX80353;
 XX AX80353;
 DT 07-SEP-1999 (first entry)

XX HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:103.

XX HCV; hepatitis C virus; single chain recombinant complex; linker;
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
 KW hydrophobic domain; covalent complex; detection; inhibitor; ss.

XX Hepatitis C virus.

OS Synthetic.

XX W09928482-A2.

XX 10-JUN-1999.

PF 24-NOV-1998; 98WO-US24528.
XX
XX 28-JUL-1998; 98US-0094331.
PR 28-NOV-1997; 97US-0067315.
XX
XX PA (SCHE) SCHERING CORP.
XX
XX MalcolM BA, Taremi SS, Weber PC, Yao N;
XX
XX WPI; 1999-385385/32.
XX
XX New hepatitis C virus covalent complexes
XX
XX Disclosure; Page 160-162; 21lpp; English.
XX
XX The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence encodes an example of the above complex. The covalent
CC NS4A-NS3 complexes are useful for structural determination and
CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.
CC They can also be used for detecting inhibitors of the protease activity,
CC the helicase activity and the ATPase activity of NS3. The covalent
CC NS4A-NS3 complexes are more soluble, stable and active than the non-
CC covalent protease-peptide complexes previously available.
XX
XX Sequence 1998 BP; 410 A; 596 C; 568 G; 424 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 1.8e-75 Length: 1998
Score: 907.50 Matches: 170
Percent Similarity: 94.39% Conservative: 15
Best Local Similarity: 86.73% Mismatches: 8
Query Match: 87.94% Indels: 3
DB: 20 Gaps: 1
US-09-965-594-16 (1-197) x AAX80353 (1-1998)
QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
DB 64 GGTCTCTGTTATTCTGTCAGTAATATTATTATCTGGTAGTGGTAGTATCATCGGCCCTAC 123
QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGlnThrSerGlnThrGlyArgAspLys 41
DB 124 TCCCAACAGACGGCGGCGCTACTTGGTTCAGAGATCACTAGGCTTACAGCGCGGACAAAG 183
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrPheLeuAlaThr 61
DB 184 AACAGGTGAGGAGAGGTTCAAGTGGTTTCACCCGCAACACATCTCTCTGGGACC 243
QY 62 CysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
DB 244 TGGGTCAACGGCGCTGTTGGACCGTTTACCATGGTGGCTGCTCAAAGACCTTAGCGCGC 303
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
DB 304 CCAAGGGGCGCAATACCCAGATGTACATAATGTGGACCGAGACCTCTCGGCTGGCAG 363
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
DB 364 GCGCCCGCGGGCGGGTCTCTTGACACCATGACCTGTGGGAGCTCAGACCTTACTITG 423
QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
DB 424 GTCAGAGACATCTGACGTCATTCGGTCCGCGCGCGGCGGACAGTAGGGGAGCCTG 483
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
DB 484 CTCTCCCGGAGGCGCTCTCTCTACTTGAAGGGGCTCTCGGGTGGTCCACTGCTGCTCCT 543
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181

DB 544 TCGGGGCACGCTGTGGCGCTCTCCGGGCTGCCGTATGCACCGGGGGTTCGGAAGCG 603
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
DB 604 GTGACTTGTGGCCGTAGAGTCATGCAACTACTATGCGGTCTCCG 651
RESULT 13
AAX80354
ID AAX80354 standard; cDNA; 1998 BP.
XX
XX AAX80354;
XX
XX 07-SEP-1999 (first entry)
XX
XX HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:104.
XX
XX HCV; hepatitis C virus; single chain recombinant complex; linker;
KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
KW hydrophobic domain; covalent complex; detection; inhibitor; ss.
XX
XX Hepatitis C virus.
OS Synthetic.
XX
XX WO9928482-A2.
XX
XX 10-JUN-1999.
XX
XX 24-NOV-1998; 98WO-US24528.
XX
XX 28-JUL-1998; 98US-0094331.
PR 28-NOV-1997; 97US-0067315.
XX
XX (SCHE) SCHERING CORP.
XX
XX MalcolM BA, Taremi SS, Weber PC, Yao N;
XX
XX WPI; 1999-385385/32.
XX
XX New hepatitis C virus covalent complexes
XX
XX Disclosure; Page 163-166; 21lpp; English.
XX
XX The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence encodes an example of the above complex. The covalent
CC NS4A-NS3 complexes are useful for structural determination and
CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.
CC They can also be used for detecting inhibitors of the protease activity,
CC the helicase activity and the ATPase activity of NS3. The covalent
CC NS4A-NS3 complexes are more soluble, stable and active than the non-
CC covalent protease-peptide complexes previously available.
XX
XX Sequence 1998 BP; 410 A; 596 C; 568 G; 424 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 1.8e-75 Length: 1998
Score: 907.50 Matches: 170
Percent Similarity: 94.39% Conservative: 15
Best Local Similarity: 86.73% Mismatches: 8
Query Match: 87.94% Indels: 3
DB: 20 Gaps: 1
US-09-965-594-16 (1-197) x AAX80354 (1-1998)
QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
DB 64 GGTCTCTGTTATTGTGGTAGAATATTATTATCTGGTAGTGGTAGTATCATCGGCCCTAC 123
QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGlnThrSerGlnThrGlyArgAspLys 41
DB 124 TCCCAACAGACGGCGGCGCTACTTGGTTCAGAGATCACTAGGCTTACAGCGCGGACAAAG 183
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrPheLeuAlaThr 61
DB 184 AACAGGTGAGGAGAGGTTCAAGTGGTTTCACCCGCAACACATCTCTCTGGGACC 243
QY 62 CysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
DB 244 TGGGTCAACGGCGCTGTTGGACCGTTTACCATGGTGGCTGCTCAAAGACCTTAGCGCGC 303
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
DB 304 CCAAGGGGCGCAATACCCAGATGTACATAATGTGGACCGAGACCTCTCGGCTGGCAG 363
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
DB 364 GCGCCCGCGGGCGGGTCTCTTGACACCATGACCTGTGGGAGCTCAGACCTTACTITG 423
QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
DB 424 GTCAGAGACATCTGACGTCATTCGGTCCGCGCGGCGGACAGTAGGGGAGCCTG 483
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
DB 484 CTCTCCCGGAGGCGCTCTCTCTACTTGAAGGGGCTCTCGGGTGGTCCACTGCTGCTCCT 543
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181

```

Db 124 TCCCAACAGACGGGGCCCTACTTGGTGCATCAAGACTAGCTTACAGCGCGGACAG 183
Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
Db 184 AACCAAGTCGAGGAGAGGTTTCAGGTGGTTTCCACCGCAACAACTCTCTCTGCGGACC 243
Qy 62 CysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
Db 244 TGGGTCAACCGCGGTGGTGGACCGGTTTACCATGGTGGCTCAAGACCTTACCGCGC 303
Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
Db 304 CCAAGAGGCGCAATCAACCCAGATGTACACTAATGTGACAGACCTCTGCGGTGGCAG 363
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
Db 364 GCGCCCGCGGGCGGCTCTTGTACACCATGACCTGTGGCAGCTTCAGACCTTTACTTG 423
Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
Db 424 GTCACGAGACATGCTGACGTCATTCGGTGGCGCGCGGGGACAGTAGGGGAGCGCTG 483
Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
Db 484 CTCCTCCCGCCAGCGCTGCTCTCTACTTGAAGGGCTCTTCGGGTGGTCCACTGCTCTGCCCT 543
Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181
Db 544 TCGGGGACGCTGTGGGCTCTTCGGGCTGCGGTATGACCCCGGGGGTTGCGAAGCGG 603
Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
Db 604 GTGACTTTGTGCCCGTAGAGTCCATGGAACTACTATGCGGTCTCCG 651

RESULT 14
ABX15706
ID ABX15706 standard; DNA; 612 BP.
XX
AC ABX15706;
XX
DT 28-MAR-2003 (first entry)
XX
DE Anti-viral synthetic prototoxophore associated DNA sequence.
XX
KW Hepatitis C; ds; viral prototoxophore; anti-viral; tumour;
KW virus; infection; antitumour; toxophore; human immunodeficiency virus;
KW HIV infection; herpes simplex virus; HSV; rhinovirus; NS3 protease.
XX
OS Unidentified.
XX
PN W0200287500-A2.
XX
PD 07-NOV-2002.
XX
PF 26-APR-2002; 2002WO-USI3223.
XX
PR 27-APR-2001; 2001US-286893P.
XX
PA (NEWB-) NEWBIOTICS INC.
XX
PI Cathers BE, Neuteboom STC, Shepard HM;
XX
DR WPI; 2003-167102/16.
XX
PT Novel synthetic viral prototoxophore for treating viral infections, has
PT toxin moiety incorporated into substrate domain specific for viral
PT enzyme, bound and modified by viral enzyme to get converted into
PT toxophore -
XX
PS Example 1; Page 62; 66pp; English.
XX
CC This invention relates to a novel synthetic viral prototoxophore

```

```

CC comprising a toxin moiety operatively incorporated into a substrate
CC domain specific for a viral enzyme. This prototoxophore may be bound
CC and modified by the viral enzyme thus converting it to a toxophore.
CC Also disclosed in the invention is a method for enhancing the anti-viral
CC effect of an antiviral agent, this method comprises contacting a cell,
CC infected with a virus or is susceptible to infection, with a
CC prototoxophore. The invention further comprises an assay to identify
CC anti-viral agents, comprising contacting an infected cell with a
CC candidate agent and comparing the ability of the agent to inhibit the
CC growth or infectivity of the virus in the cell. The prototoxophores
CC of the invention may have virucide or antitumour activity. The
CC prototoxophores of the invention may be useful for reducing or
CC inhibiting viral infectivity, by contacting a cell (e.g. lymphocyte,
CC nerve cell, connective tissue cell, muscle cell or hepatocyte) which is
CC infected with a virus or is susceptible to infection with a virus, with
CC an effective amount of the prototoxophore. The cells are cell lines
CC adapted to long term continuous culture or isolated from a subject.
CC The prototoxophore is also useful for ameliorating the severity of a
CC viral infection in a subject, where the virus is selected from human
CC immunodeficiency virus (HIV), herpes simplex virus (HSV), rhinovirus and
CC hepatitis virus, by administering an effective amount of the
CC prototoxophore to the subject. The prototoxophores of the invention are
CC also useful for treating tumours. The present sequence represents an
CC antiviral prototoxophore associated DNA sequence, this sequence is
CC described as a recombinant NS3/NS4 fusion protein in example 1 of
CC the invention although it is clearly not a protein sequence.
XX
SQ Sequence 612 BP; 120 A; 171 C; 191 G; 130 T; 0 other;

```

```

Alignment Scores:
Pred. No.: 7.89e-76 Length: 612
Score: 904.50 Matches: 178
Percent Similarity: 92.82% Conservative: 3
Best Local Similarity: 91.28% Mismatches: 11
Query Match: 87.65% Indels: 3
DB: 25 Gaps: 1

US-09-965-594-16 (1-197) x ABX15706 (1-612)

Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
Db 19 GGTAGTGTGGTGCATTTGGGTAGGATCATTTGCGGTAGTGGTAGTATCATCGCGGTAC 78
Qy 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41
Db 79 GCCCAGCAGACAAGGGGCTCTCTAGGGTGCATATCACCAGGCTAACTGCCCGGGACAAA 138
Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
Db 139 AACCAAGTGGGGTGGAGTCCAGATGTGTCACTGTGCTGCCCAACCTTCTCTGGCAAG 198
Qy 62 CysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
Db 199 TGCATCAATGGGTGTGCTGGACTGTCTACCGGGCGGCGGAACGAGGACCATCGCGTCA 258
Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
Db 259 CCCAAGGTCTCTGTCAATCCAGATGTATACCAATGTAGACCAAGACCTTGTGGGCTGGGCC 318
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
Db 319 GCTTCGCAAGGTACCGGCTCATTTGACACCCCTGCGCTTGGGCTCTCGGACCTTTACCTG 378
Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
Db 379 GTCACGAGGACCGCGATGTCTCCCGTGGCGCGGGGGTGATAGCAGGGGCGACGCTG 438
Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
Db 439 CTGTGCGCGCGCGCCCATTTCTTGAAGGGCTCTCTCGGGGGGGTCCGCTGTGTGCGCCC 498
Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 30, 2003, 19:26:03 ; Search time 176.482 Seconds
(without alignments)
2560.981 Million cell updates/sec

Title: US-09-965-594-16

Perfect score: 1032

Sequence: 1 MKKGSVVIVGRINLSGDTA.....VAKAVDFIPVESLETTMRSP 197

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1533700 seqs, 1147125425 residues

Total number of hits satisfying chosen parameters: 3067400

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n.model -DEV=xlp
-O=/cgn2_1/USPTO_spool/US09965594/runat_29082003_151920_28367/app.query.fasta_1.2872
-DB=PublishedApplications_NA -OFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=BL0SUM62
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09965594 @cgn_1.1.864 @runat_29082003_151920_28367
-NCPU=6 -ICPU=3 -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_NA.*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1032	100.0	594	10	US-09-965-594-17 Sequence 17, Appl

2	1015	98.4	594	10	US-09-965-594-15	Sequence 15, Appl
3	1002	97.1	594	10	US-09-965-594-19	Sequence 19, Appl
4	990	95.9	594	10	US-09-965-594-21	Sequence 21, Appl
5	981	95.1	588	10	US-09-965-594-13	Sequence 13, Appl
6	980	95.0	594	10	US-09-965-594-23	Sequence 23, Appl
7	976	94.6	594	10	US-09-965-594-25	Sequence 25, Appl
8	942	91.3	588	10	US-09-965-594-4	Sequence 4, Appl
9	904.5	87.6	612	14	US-10-133-133A-6	Sequence 6, Appl
10	888.5	86.1	9846	9	US-09-742-653-3	Sequence 3, Appl
11	888.5	86.1	9846	10	US-09-238-076-1	Sequence 1, Appl
12	888.5	86.1	9846	11	US-09-995-937-1	Sequence 1, Appl
13	888.5	86.1	9846	11	US-09-917-563-1	Sequence 1, Appl
14	888.5	86.1	12980	10	US-09-238-076-5	Sequence 5, Appl
15	888.5	86.1	12980	11	US-09-995-937-5	Sequence 5, Appl
16	888.5	86.1	12980	11	US-09-917-563-5	Sequence 5, Appl
17	884.5	85.7	9379	9	US-09-916-359-1	Sequence 1, Appl
18	884.5	85.7	9416	10	US-09-238-076-19	Sequence 19, Appl
19	884.5	85.7	9416	11	US-09-995-937-19	Sequence 19, Appl
20	884.5	85.7	9416	11	US-09-917-563-19	Sequence 19, Appl
21	882	85.5	549	10	US-09-965-594-2	Sequence 2, Appl
22	882	85.5	2058	10	US-09-881-654-1	Sequence 1, Appl
23	882	85.5	2058	10	US-09-881-239-2	Sequence 2, Appl
24	881.5	85.4	836	10	US-09-921-337-120	Sequence 120, App
25	881.5	85.4	10803	10	US-09-747-419-17	Sequence 17, Appl
26	881.5	85.4	10803	14	US-10-259-275-17	Sequence 17, Appl
27	878.5	85.1	9416	10	US-09-929-955-13	Sequence 13, Appl
28	878.5	85.1	9416	13	US-10-104-966-13	Sequence 13, Appl
29	878	85.1	2061	10	US-09-929-955-16	Sequence 16, Appl
30	866.5	84.0	13910	11	US-09-919-901-1	Sequence 1, Appl
31	863.5	83.7	13910	11	US-09-919-901-8	Sequence 8, Appl
32	863.5	83.7	13910	11	US-09-919-901-15	Sequence 15, Appl
33	863	83.6	2064	11	US-09-884-456-69	Sequence 69, Appl
34	863	83.6	2523	11	US-09-884-456-85	Sequence 85, Appl
35	860.5	83.4	2073	14	US-10-131-133A-5	Sequence 5, Appl
36	860	83.3	6189	14	US-10-259-275-41	Sequence 41, Appl
37	860	83.3	7992	13	US-10-005-469-1	Sequence 1, Appl
38	860	83.3	7992	13	US-10-005-469-2	Sequence 2, Appl
39	860	83.3	7992	13	US-10-005-469-4	Sequence 4, Appl
40	860	83.3	7992	13	US-10-005-469-6	Sequence 6, Appl
41	860	83.3	8638	12	US-10-309-561-24	Sequence 24, Appl
42	860	83.3	8638	13	US-10-029-907-24	Sequence 24, Appl
43	860	83.3	8639	12	US-10-309-561-1	Sequence 1, Appl
44	860	83.3	8639	13	US-10-029-907-1	Sequence 1, Appl
45	860	83.3	8642	12	US-10-309-561-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-09-965-594-17
; Sequence 17, Application US/09965594
; Patent No. US20020106642A1
; GENERAL INFORMATION:
; APPLICANT: Wittekind, Michael
; APPLICANT: Weinheimer, Steven
; APPLICANT: Zhang, Yaqun
; APPLICANT: Goldfarb, Valentina
; TITLE OF INVENTION: Modified Forms of Hepatitis C NS3 Protease for Facilitating Inhibitor Screening and Structural Studies
; TITLE OF INVENTION: Facilitating Inhibitor Screening and Structural Studies
; FILE REFERENCE: DB17Sequences
; CURRENT APPLICATION NUMBER: US/09/965,594
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/115,271
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Hepatitis C virus
US-09-965-594-17

•

; TYPE: DNA
; ORGANISM: Hepatitis C virus
US-09-965-594-19

Alignment Scores:
Pred. No.: 5,45e-107 Length: 594
Score: 1002.00 Matches: 194
Percent Similarity: 98.48% Conservatives: 0
Best Local Similarity: 98.48% Mismatches: 3
Query Match: 97.09% Indels: 0
DB: 10 Gaps: 0

US-09-965-594-16 (1-197) x US-09-965-594-19 (1-594)

```
QY 1 MetLysLysLysGlySerValValleValGlyArgIleAsnLeuSerGlyAspThrAla 20
DB 1 ATGAAAAAAGGATCGTGTGTTATCGTCGGCGTATCAACCTGTCGGGTGACACCGCT 60
QY 21 TyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAsp 40
DB 61 TACGCTCAGCAGACTCGAGGTGAGAGGGTGCAGAAACCTCCAGACGGGTGCTGAC 120
QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
DB 121 AAAAACCAAGTTGAAGTGAAGTTGAGATCGTTTCCACCGCTACCCAGACCTTCCTGGCT 180
QY 61 ThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
DB 181 ACCTCCATCAACGGTGTCTGTGACCGCTTACCACCGGTGCTGATACCCGTACCATCGCT 240
QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr 100
DB 241 TCCCGGAAAGTCCCGTTACCCAGATGTACACCAACGTTGACAAAGACCTGGTGGTGG 300
QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
DB 301 CAGGCTCCGAGGGTCCCGTCCCTGACCGGTGACCGCTGCTGCTCCCGACCTGTAC 360
QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
DB 361 CTGGTTACCCGTCAGCTGACGTTATCCCGGTTCGTGCTGGTGACTCCCGTGGTTC 420
QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
DB 421 CTGCTGTCCCGCGTCCGATCTCCTACCTGAAAGGTTCTCCCGTGGTCCGCTGCTGC 480
QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys 180
DB 481 CCGGCTGGTCAAGCTGTGGTATCTCCGTGCTGCTGTTCCACCGGTGGTGGTAA 540
QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
DB 541 GCTGTTGACTTCATCCCGGTTGAATCCCTGGAAACCAACCATGCGTTCCTCCCG 591
```

RESULT 4

US-09-965-594-21

; Sequence 21, Application US/09965594

; Patent No. US20020106642

; GENERAL INFORMATION:

; APPLICANT: Wittekind, Michael

; APPLICANT: Weinheimer, Steven

; APPLICANT: Zhang, Yaquin

; APPLICANT: Goldfarb, Valentina

; TITLE OF INVENTION: Modified Forms of Hepatitis C NS3 Protease for
; TITLE OF INVENTION: Facilitating Inhibitor Screening and Structural Studies

; FILE OF INVENTION: of Protease: Inhibitor Complexes

; FILE REFERENCE: DB17Sequences

; CURRENT APPLICATION NUMBER: US/09/965,594

; PRIOR FILING DATE: 2001-09-27

; PRIOR FILING DATE: 1999-01-08

; NUMBER OF SEQ ID NOS: 26

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 21
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Hepatitis C virus
US-09-965-594-21

Alignment Scores:
Pred. No.: 1.34e-105 Length: 594
Score: 990.00 Matches: 191
Percent Similarity: 97.97% Conservatives: 2
Best Local Similarity: 96.95% Mismatches: 4
Query Match: 95.93% Indels: 0
DB: 10 Gaps: 0

US-09-965-594-16 (1-197) x US-09-965-594-21 (1-594)

```
QY 1 MetLysLysLysGlySerValValleValGlyArgIleAsnLeuSerGlyAspThrAla 20
DB 1 ATGAAAAAAGGATCGTGTGTTATCGTCGGCGTATCAACCTGTCGGGTGACACCGCT 60
QY 21 TyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAsp 40
DB 61 TACGCTCAGCAGACTCGAGGTGAGAGGGTGCAGAAAGACCTCCACACCGGTGCTGAC 120
QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
DB 121 AAAAACCAAGTTGAAGTGAAGTTGAGATCGTTTCCACCGCTACCCAGACCTTCCTGGCT 180
QY 61 ThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
DB 181 ACCTCCATCAACGGTGTCTGTGACCGCTTACCACCGGTGCTGATACCCGTACCATCGCT 240
QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr 100
DB 241 TCCCGGAAAGTCCCGTTACCCAGATGTACACCAACGTTGACAAAGACCTGGTGGTGG 300
QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
DB 301 CAGGCTCCGAGGGTCCCGTCCCTGACCGGTGACCGCTGCTGCTCCCGACCTGTAC 360
QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
DB 361 CTGGTTACCCGTCAGCTGACGTTATCCCGGTTCGTGCTGGTGACTCCCGTGGTTC 420
QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
DB 421 CTGCTGTCCCGCGTCCGATCTCCTACCTGAAAGGTTCTCCCGTGGTCCGCTGCTGC 480
QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys 180
DB 481 CCGGCTGGTCAAGCTGTGGTATCTCCGTGCTGCTGTTCCACCGGTGGTGGTAA 540
QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
DB 541 GCTGTTGACTTCATCCCGGTTGAATCCCTGGAAACCAACCATGCGTTCCTCCCG 591
```

RESULT 5

US-09-965-594-13

; Sequence 13, Application US/09965594

; Patent No. US20020106642

; GENERAL INFORMATION:

; APPLICANT: Wittekind, Michael

; APPLICANT: Weinheimer, Steven

; APPLICANT: Zhang, Yaquin

; APPLICANT: Goldfarb, Valentina

; TITLE OF INVENTION: Modified Forms of Hepatitis C NS3 Protease for

; TITLE OF INVENTION: Facilitating Inhibitor Screening and Structural Studies

; FILE OF INVENTION: of Protease: Inhibitor Complexes

; FILE REFERENCE: DB17Sequences

; CURRENT APPLICATION NUMBER: US/09/965,594

; PRIOR FILING DATE: 2001-09-27

; PRIOR FILING DATE: 1999-01-08

; NUMBER OF SEQ ID NOS: 26

; SOFTWARE: PatentIn Ver. 2.0

```

; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 13
; LENGTH: 588
; TYPE: DNA
; ORGANISM: Hepatitis C virus
US-09-965-594-13

Alignment Scores:
Pred. No.: 1,46e-104 Length: 588
Score: 981.00 Matches: 190
Percent Similarity: 96.95% Conservative: 1
Best Local Similarity: 96.45% Mismatches: 4
Query Match: 95.08% Indels: 2
DB: 1 Gaps: 1

US-09-965-594-16 (1-197) x US-09-965-594-13 (1-588)
QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
DB 1 ATGAAAAAAGATCCGTTGTTATCGTCGGCGGTATAGTACTGACCGGT-----GCT 54
QY 21 TyrAlaGlnGlnThrArgGlyGluGlyCysGlnGlnThrSerGlnThrGlyArgAsp 40
DB 55 TAGCGTCACGACACTCGAGTTCGAGGAGGTGCCAAGAAACCTCCACAGCCGGTCGTGAC 114
QY 41 LysAsnGlnValGluGlyClyValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
DB 115 AAAAAACAGGTGAAGGTGAAGTTTCAGATCGTTCCACCGCTGCTCAGACCTTCCTGGCT 174
QY 61 ThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
DB 175 ACCTGCATCAACGGTGTTCGTCGACCGTTTACCACCGTGTGTTACCGTACCATCGCT 234
QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp 100
DB 235 TCCCCGAAGGTCCGGTTATCCAGATGTACACCAACGTTGACAAAGACCTGGTTGGTTGG 294
QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
DB 295 CCGGCTCCGACGGTTCCTCCCTGACCGCGTTCACCGTCCGCGTTCCTCCGACCTGTAC 354
QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
DB 355 CTGGTTACCCGTCACGCTGACGTTATCCCGGTTCTGCTGCTGCTGCTGCTGCTGCTGCT 414
QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
DB 415 CTGCTGTCCCGGCTCCGATCTCTACCTGAAAGGTTCCTCCGGTGGTCCGCTGCTGTC 474
QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys 180
DB 475 CCGGCTGGTCACGCTGTGGTATCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTAAA 534
QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
DB 535 GCCTGTGACTTCATCCCGGTTGAAATCCCTGGAAACCAACCATGCGTTCCTCCCG 585

RESULT 6
US-09-965-594-23
; Sequence 23, Application US/09965594
; Patent No. US20020106642A1
; GENERAL INFORMATION:
; APPLICANT: Wittekind, Michael
; APPLICANT: Weinheimer, Steven
; APPLICANT: Zhang, Yaqu
; APPLICANT: Goldfarb, Valentina
; TITLE OF INVENTION: Modified Forms of Hepatitis C NS3 Protease for
; TITLE OF INVENTION: Facilitating Inhibitor Screening and Structural Studies
; TITLE OF INVENTION: of Protease: Inhibitor Complexes
; FILE REFERENCE: DB17Sequences
; CURRENT APPLICATION NUMBER: US/09/965,594
; CURRENT FILING DATE: 2001-09-27

; PRIOR APPLICATION NUMBER: 60/1115,271
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 23
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Hepatitis C virus
US-09-965-594-23

Alignment Scores:
Pred. No.: 1,93e-104 Length: 594
Score: 980.00 Matches: 190
Percent Similarity: 97.46% Conservative: 2
Best Local Similarity: 96.45% Mismatches: 5
Query Match: 94.96% Indels: 0
DB: 1 Gaps: 0

US-09-965-594-16 (1-197) x US-09-965-594-23 (1-594)
QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
DB 1 ATGAAAAAAGATCCGTTGTTATCGTCGGCGGTATCAACCTGTCGGGTGACACCGCT 60
QY 21 TyrAlaGlnGlnThrArgGlyGluGlyCysGlnGlnThrSerGlnThrGlyArgAsp 40
DB 61 TAGCGTCACGACACTCGAGTTCGAGGAGTACCCAGAAGACCTCCACACCGGTCGTGAC 120
QY 41 LysAsnGlnValGluGlyClyValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
DB 121 AAAAAACAGGTGAAGGTGAAGTTTCAGATCGTTTCCACCGCTACCCAGACCTTCCCTGGCT 180
QY 61 ThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
DB 181 ACCTCCATCAACGGTGTTCCTGTCGACCGGTTTACCACCGTGTGTTACCGTACCATCGCT 240
QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp 100
DB 241 TCCCCGAAGGTCCGGTTTACCCAGATGTACACCAACGTTGACAAAGACCTGGTTGGTTGG 300
QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
DB 301 CAGGCTCCGACGGTTCCTGTCGACCGCGTGCACCTCGGTTCTCCGACCTGTAC 360
QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
DB 361 CTGGTTACCCGTCACGCTGACGTTATCCCGGTTCTGCTGCTGCTGCTGCTGCTGCTGCT 420
QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
DB 421 CTGCTGTCCCGGCTCCGATCTCTACCTGAAAGGTTCCTCCGGTGGTCCGCTGCTGTC 480
QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys 180
DB 481 CCGGCTGGTCACGCTGTGGTATCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTAAA 540
QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
DB 541 GCTGTTGACTTCATCCCGGTTGAAATCCCTGGAAACCAACCATGCGTTCCTCCCG 591

RESULT 7
US-09-965-594-25
; Sequence 25, Application US/09965594
; Patent No. US20020106642A1
; GENERAL INFORMATION:
; APPLICANT: Wittekind, Michael
; APPLICANT: Weinheimer, Steven
; APPLICANT: Zhang, Yaqu
; APPLICANT: Goldfarb, Valentina
; TITLE OF INVENTION: Modified Forms of Hepatitis C NS3 Protease for
; TITLE OF INVENTION: Facilitating Inhibitor Screening and Structural Studies
; TITLE OF INVENTION: of Protease: Inhibitor Complexes
; FILE REFERENCE: DB17Sequences
; CURRENT APPLICATION NUMBER: US/09/965,594
; CURRENT FILING DATE: 2001-09-27

```

; CURRENT APPLICATION NUMBER: US/09/965,594
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/115,271
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Hepatitis C virus
US-09-965-594-25

Alignment Scores:
Pred. No.: 5,62e-104 Length: 594
Score: 976.00 Matches: 189
Percent Similarity: 95.94% Conservative: 0
Best Local Similarity: 95.94% Mismatches: 8
Query Match: 94.57% Indels: 0
DB: 10 Gaps: 0

US-09-965-594-16 (1-197) x US-09-965-594-25 (1-594)

```
QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
DB 1 ATCAAAAAAAGGATCGTTGTTATCGTCGGCGGTATCAACCTGTCGGGTGACCGCT 60
QY 21 TyrAlaGlnGlnThrArgGlyGluGlyCysGlnGlnThrSerGlnThrGlyArgAsp 40
DB 61 TACGCTCAGCAGACTCGAGGTCCTGGTGTGATCATCCTCCCTGACCGGTGTCGAC 120
QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
DB 121 AAAAACAGGTTGAAGTGAAGTTCAGATCGTTCCACCGCTGTCAGACCTCTCTGGCT 180
QY 61 ThrCysIleAsnGlyValCysTrpThrValTyHisGlyAlaGlyThrArgThrIleAla 80
DB 181 ACCTGCATCAACCGGTGTTGCTGACCGCTTTACCACCGTCTGCTACCCGTACCATCGCT 240
QY 81 SerProLysGlyProValThrGlnMetTyThrAsnValAspLysAspLeuValGlyTrp 100
DB 241 TCCCGGAAGGTCGGTATCCAGATGATACACCAAGCTTGACAAGACCTGGTGGTGG 300
QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTy 120
DB 301 CCGGCTCGCAGGTTCCGCTTCCTGACCGCGTGCACCTCGGTTCTCCGACCTGTAC 360
QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
DB 361 CTGGTTACCGGTACGCTGACGTTATCCCGGTTCTGCTGCTGCTGCTGCTGCTGCTG 420
QY 141 LeuLeuSerProArgProIleSerTyrtLeuLysGlySerSerGlyGlyProLeuLeu 160
DB 421 CTGCTGTCCCGCGCTCCGATCTCCTACCTGAAGGTTCTCCCGGTGGTCCGCTGTC 480
QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys 180
DB 481 CCGGCTGGTCAGCTGTGGTATCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTAAA 540
QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
DB 541 GCTGTGTACTTCATCCCGGTTGAATCCCTGGAAACCAACCATGCTTCCCGC 591
```

RESULT 8

US-09-965-594-4
; Sequence 4, Application US/09965594
; Patent No. US20020106642A1
; GENERAL INFORMATION:
; APPLICANT: Wittekind, Michael
; APPLICANT: Weinheimer, Steven
; APPLICANT: Zhang, Yaqun
; APPLICANT: Goldfarb, Valentina
; TITLE OF INVENTION: Modified Forms of Hepatitis C NS3 Protease for
; Facilitating Inhibitor Screening and Structural Studies

; TITLE OF INVENTION: of Protease: Inhibitor Complexes
; FILE REFERENCE: DB17Sequences
; CURRENT APPLICATION NUMBER: US/09/965,594
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/115,271
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 588
; TYPE: DNA
; ORGANISM: Hepatitis C virus
US-09-965-594-4

Alignment Scores:
Pred. No.: 4,84e-100 Length: 588
Score: 942.00 Matches: 185
Percent Similarity: 94.42% Conservative: 1
Best Local Similarity: 93.91% Mismatches: 9
Query Match: 91.28% Indels: 2
DB: 10 Gaps: 1

US-09-965-594-16 (1-197) x US-09-965-594-4 (1-588)

```
QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
DB 1 ATCAAAAAAAGGTTCCGTTGTTATCGTCGGCGGTATAGTACTGAACGGT-----GCT 54
QY 21 TyrAlaGlnGlnThrArgGlyGluGlyCysGlnGlnThrSerGlnThrGlyArgAsp 40
DB 55 TACGCTCAGCAGACTCGAGGTCCTGGTGTGTCATCATCCTCCCTGACCGGTGTCGAC 114
QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
DB 115 AAAAACAGGTTGAAGTGAAGTTCAGATCGTTCCACCGCTGTCAGACCTCTCTGGCT 174
QY 61 ThrCysIleAsnGlyValCysTrpThrValTyHisGlyAlaGlyThrArgThrIleAla 80
DB 175 ACCTGCATCAACCGGTGTTGCTGACCGCTTTACCACCGTGTGTTACCGTACCATCGCT 234
QY 81 SerProLysGlyProValThrGlnMetTyThrAsnValAspLysAspLeuValGlyTrp 100
DB 235 TCCCGGAAGGTCGGTATCCAGATGATACACCAAGCTTGACAAGACCTGGTGGTGG 294
QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTy 120
DB 295 CCGGCTCGCAGGTTCCGTTCCCTGACCGCGTGCACCTCGGTTCTCCGACCTGTAC 354
QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
DB 355 CTGGTTACCGGTACGCTGACGTTATCCCGGTTCTGCTGCTGCTGCTGCTGCTGCTTC 414
QY 141 LeuLeuSerProArgProIleSerTyrtLeuLysGlySerSerGlyGlyProLeuLeu 160
DB 415 CTGCTGTCCCGCGCTCCGATCTCCTACCTGAAGGTTCTCCCGGTGGTCCGCTGCTG 474
QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys 180
DB 475 CCGGCTGGTCAGCTGTGGTATCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTAAA 534
QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
DB 535 GCTGTGTACTTCATCCCGGTTGAATCCCTGGAAACCAACCATGCTTCCCGC 585
```

RESULT 9

US-10-133-133A-6
; Sequence 6, Application US/10133133A
; Publication No. US20030114385A1
; GENERAL INFORMATION:
; APPLICANT: CATHERS, Brian
; APPLICANT: NEUTEBOON, Saskia
; APPLICANT: SHEPARD, Michael
; TITLE OF INVENTION: VIRAL ENZYME ACTIVATED PROTOXOPHORES

; TITLE OF INVENTION: AND USE OF SAME TO TREAT VIRAL INFECTIONS
; FILE REFERENCE: NB 2021.00
; CURRENT APPLICATION NUMBER: US/10/133,133A
; CURRENT FILING DATE: 2002-04-26
; PRIOR APPLICATION NUMBER: 60/286,983
; PRIOR FILING DATE: 2001-04-27
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 612
; TYPE: DNA
; ORGANISM: Hepatitis C. Virus
US-10-133-133A-6

Alignment Scores:
Pred. No.: 1,13e-95 Length: 612
Score: 904.50 Matches: 178
Percent Similarity: 92.82% Conservative: 3
Best Local Similarity: 91.28% Mismatches: 11
Query Match: 87.65% Indels: 3
DB: 14 Gaps: 1

US-09-965-594-16 (1-197) x US-10-133-133A-6 (1-612)

```
Qy 5 GlySerValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
Db 19 GSTAGTGTGTCATTGGGTAGGATCATTTTCGGTAGTGTAGTATCATCGCGGTAC 78
Qy 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41
Db 79 GCCACGACAGACAGGGGCTCTTGGGTGCATATACACAGGCTAACTGGCGCGGACAAA 138
Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
Db 139 AACCAAGTCAGGGGTGAGGTCTCAGATGTGTCACTGCTGCCCAAAACCTTCCTGGCAACG 198
Qy 62 CysIleAsnGlyValCysTrpThrValThrHisGlyAlaGlyThrArgThrIleAlaSer 81
Db 199 TGCATCAATGGGGTGTCTGAGATGTCTACCGGGGCGGACAGGACCATCGCGTCA 258
Qy 82 ProLysGlyProValThrGlnMetThrAsnValAspLysAspLeuValGlyTrpGln 101
Db 259 CCCAAGGTCCTGTCCATCCAGATGTATACCAATGTAGACCAAGACCTTGTGGCTGGCCC 318
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuThrLeu 121
Db 319 GCTTCCGAAGGTACCGGCTCATTCACACCTGCACTTGGGCTCTCTCGGACCTTTACCTG 378
Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
Db 379 GTCACGAGGACCGCGATGTATCCGTGCGCGGGGGGTGATAGCAGGGGACGCTG 438
Qy 142 LeuSerProArgProIleSerThrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
Db 439 CTGTGCGCGCGGGCCCAATTCCTACTTTGAAAGGCTCTCGGGGGGTTCGCTGTGTGCCCC 498
Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181
Db 499 GCGGGGACCGCGGTGGGCATATTATAGGCGCGCGTGTGCACCGGTGAGTGGCTAAGGCG 558
Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196
Db 559 GTGACATTTATCCCTGTGGAGACCTTAGACACAACCATGAGGTCC 603
```

RESULT 10

US-09-742-659-3
; Sequence 3, Application US/09742659
; Patent No. US20010034019A1
GENERAL INFORMATION:
; APPLICANT: Hong, Zhi
; APPLICANT: Butkiewicz, Nancy J.
; APPLICANT: Zhong, Weidong
; APPLICANT: Ingravallo, Paul

; APPLICANT: Wright-Minogue, Jacquelyn
; APPLICANT: Lau, Johnson Y.
; APPLICANT: Lemon, Stanley M.
; TITLE OF INVENTION: Chimeric HCV/GBV-B viruses
; FILE REFERENCE: ID01116
; CURRENT APPLICATION NUMBER: US/09/742,659
; CURRENT FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: US 60/171,469
; PRIOR FILING DATE: 1999-12-22
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 9646
; TYPE: DNA
; ORGANISM: Hepatitis C virus
US-09-742-659-3

Alignment Scores:
Pred. No.: 2,78e-92 Length: 9646
Score: 888.50 Matches: 172
Percent Similarity: 89.22% Conservative: 10
Best Local Similarity: 84.31% Mismatches: 13
Query Match: 86.09% Indels: 9
DB: 1 Gaps: 1

US-09-965-594-16 (1-197) x US-09-742-659-3 (1-9646)

```
Qy 3 LysLysGlySerValIleValGlyArgIleAsn----- 14
Db 3354 CGTAGGGCCGAGGATGATCTGTGGACACCGCGGATGCTCCCAAGGGGTGGAGG 3413
Qy 15 ---LeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGlyCysGlnGlu 33
Db 3414 TTGCTGTGCGCCCATCAGCGGTACGCCGACGACGAGAGGCGCTCTCTGGTGTATAATC 3473
Qy 34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyValGlnIleValSerThr 53
Db 3474 ACCAGTCTGACTGGCGGGACAAAACCAAGTGGAGGTGAGTCCGATCGTGTCACT 3533
Qy 54 AlaThrGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValThrHisGly 73
Db 3534 GCTACCCAAACCTCTCTGGCAACGTGCATCAATGGGTATGCTGGACTGCTACCAAGGG 3593
Qy 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValThrGlnMetThrThrAsnVal 93
Db 3594 GCGGGAACGAGGACCATCGCATCACCACCAAGGGTCTCTATCCAGATGTATACCAATGTG 3653
Qy 94 AspLysAspLeuValGlyTrpGlnAlaProGlnGlySerArgSerLeuThrProCysThr 113
Db 3654 GACCAAGACCTTGTGGGTGGCGGCTCTCTCAAGGTTCGCGCTCATGACACCTGCACC 3713
Qy 114 CysGlySerSerAspLeuThrLeuValThrArgHisAlaAspValIleProValArgArg 133
Db 3714 TGCGGCTCTCTCGGACCTTTTACCTGGTTACGAGGACGCGGACGCTATTCGCGTGGCGG 3773
Qy 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerThrLeuLysGlySer 153
Db 3774 CGAGGTGATAGCAGGGGTAGCTGCTTTCCCGCGGCGCCATTTCTACCTAAAGGCTCC 3833
Qy 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 173
Db 3834 TCGGGGGTCCGCTGTGTGTCGCGCGGACGACGCGCTATTTCAGGGCGCGGGTG 3893
Qy 174 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 193
Db 3894 TGCACCGGTGGAGTGGCCAAAGCGGTGAGCTTTATCCCTGTGGAGAACCTAGACACAAC 3953
Qy 194 MetArgSerPro 197
Db 3954 ATGAGATCCCCG 3965
```

RESULT 11

US-09-238-076-1

Alignment Scores:

```

Pred. No.: 2,78e-92 Length: 9646
Score: 888.50 Matches: 172
Percent Similarity: 89.22% Conservative: 10
Best Local Similarity: 84.31% Mismatches: 13
Query Match: 86.09% Indels: 9
DB: 11 Gaps: 1

US-09-965-594-16 (1-197) x US-09-995-937-1 (1-9646)

QY 3 LysLysGlySerValValIleValGlyArgIleAsn----- 14
DB 3354 CGTAGGGCCAGGAGATACCTGCTGGCCACGCGGAAATGCTCTCCAAAGGGGTGGAGG 3413
QY 15 --LeuSerGlyAspThrAlaTyraAlaGlnGlnThrArgGlyGluGlnGlu 33
DB 3414 TTGCTGGCGCCATCAGCGGCTAGCCCGGACGAGAGAGGCTCTAGGGTGTATAATC 3473
QY 34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlnValGlnIleValSerThr 53
DB 3474 ACCAGCCTGACTGGCGGGACAAAACCAAGTGGAGGTGAGGTCCAGATCGTCAACT 3533
QY 54 AlaThrGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyHISGly 73
DB 3534 GCTACCCAAACCTTCTGGCAACGTGCTCAATGGGGTATGCTGGACTGTCTACACGGG 3593
QY 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValThrGlnMetTyThrAsnVal 93
DB 3594 GCGGACGAGGACCATCGCATCCCAAGGGTCTGTCTATCCAGATGTATACCAATGTG 3653
QY 94 AspLysAspLeuValGlyTrpGlnAlaProGlnGlySerArgSerLeuThrProCysThr 113
DB 3654 GACCAAGACCTTGTGGCTGGCGCTCCCAAGTTCCCGCTCATTTGACCCCTGCACC 3713
QY 114 CysGlySerSerAspLeuTyLeuValThrArgHisAlaAspValIleProValArgArg 133
DB 3714 TGGCGCTCTCGGACCTTACCTGGTTCACGAGCAGCGCGATCTATCCGTGGCGCGG 3773
QY 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyLeuLysGlySer 153
DB 3774 CGAGGTGATAGCAGGGGTAGCTGCTTTCGCCCGCGCCCATTTCTACTTGAAGGCTCC 3833
QY 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 173
DB 3834 TCGGGGGTCCGCTGTGTGTCGCCCGGGACACCGCTGGAGAACCTTAGAGCAACC 3953
QY 194 MetArgSerPro 197
DB 3954 ATGAGATCCCGC 3965

RESULT 13
US-09-917-563-1
; Sequence 1, Application US/09917563
; Publication No. US20030073080A1
; GENERAL INFORMATION:
; APPLICANT: RICE, CHARLES et al.
; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
; VIRUS (HCV) AND USES THEREOF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAFERKAMP, L.C.
; STREET: 7733 FORSYTH BLVD., SUITE 1400
; CITY: ST. LOUIS
; STATE: MO
; COUNTRY: USA
; ZIP: 63105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/917,563
; FILING DATE: 27-Jul-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/238,076
; FILING DATE: 26-JAN-1999
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 6029-4831
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 314-727-5188
; TELEFAX: 314-727-6092
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9646 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-917-563-1

Alignment Scores:
Pred. No.: 2,78e-92 Length: 9646
Score: 888.50 Matches: 172
Percent Similarity: 89.22% Conservative: 10
Best Local Similarity: 84.31% Mismatches: 13
Query Match: 86.09% Indels: 9
DB: 11 Gaps: 1

US-09-965-594-16 (1-197) x US-09-917-563-1 (1-9646)

QY 3 LysLysGlySerValValIleValGlyArgIleAsn----- 14
DB 3354 CGTAGGGCCAGGAGATACCTGCTGGCCACGCGGAAATGCTCTCCAAAGGGGTGGAGG 3413
QY 15 --LeuSerGlyAspThrAlaTyraAlaGlnGlnThrArgGlyGluGlnGlu 33
DB 3414 TTGCTGGCGCCATCAGCGGCTAGCCCGGACGAGAGAGGCTCTAGGGTGTATAATC 3473
QY 34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlnValGlnIleValSerThr 53
DB 3474 ACCAGCCTGACTGGCGGGACAAAACCAAGTGGAGGTGAGGTCCAGATCGTCAACT 3533
QY 54 AlaThrGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyHISGly 73
DB 3534 GCTACCCAAACCTTCTGGCAACGTGCTCAATGGGGTATGCTGGACTGTCTACACGGG 3593
QY 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValThrGlnMetTyThrAsnVal 93
DB 3594 GCGGACGAGGACCATCGCATCCCAAGGGTCTGTCTATCCAGATGTATACCAATGTG 3653
QY 94 AspLysAspLeuValGlyTrpGlnAlaProGlnGlySerArgSerLeuThrProCysThr 113
DB 3654 GACCAAGACCTTGTGGCTGGCGCTCCCAAGTTCCCGCTCATTTGACCCCTGCACC 3713
QY 114 CysGlySerSerAspLeuTyLeuValThrArgHisAlaAspValIleProValArgArg 133
DB 3714 TGGCGCTCTCGGACCTTACCTGGTTCACGAGCAGCGCGATCTATCCGTGGCGCGG 3773
QY 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyLeuLysGlySer 153
DB 3774 CGAGGTGATAGCAGGGGTAGCTGCTTTCGCCCGCGCCCATTTCTACTTGAAGGCTCC 3833
QY 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 173
DB 3834 TCGGGGGTCCGCTGTGTGTCGCCCGGGACACCGCTGGAGAACCTTAGAGCAACC 3953

```


QY 174 CysThrArgGlyValAlaValAlaValAspPheIleProValGluSerLeuGluThr 193
Db 3894 TGCACCGTGGAGTGGCTAAGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACACC 3953
QY 194 MetArgSerPro 197
Db 3954 ATGAGATCCCG 3965
RESULT 14
US-09-238-076-5
; Sequence 5, Application US/09238076
; Patent No. US20020102540A1
; GENERAL INFORMATION:
; APPLICANT: RICE, CHARLES et al.
; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAERKAMP, L.C.
; STREET: 7733 FORSYTH BLVD., SUITE 1400
; CITY: ST. LOUIS
; STATE: MO
; COUNTRY: USA
; ZIP: 63105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/238.076
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/034,756
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 6029-4831
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 314-727-5188
; TELEFAX: 314-727-6092
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12980 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-09-238-076-5
Alignment Scores:
Pred. No.: 4.06e-92 Length: 12980
Score: 888.50 Matches: 172
Percent Similarity: 89.22% Conservative: 10
Best Local Similarity: 84.31% Mismatches: 13
Query Match: 86.09% Indels: 9
DB: 10 Gaps: 1
US-09-965-594-16 (1-197) x US-09-238-076-5 (1-12980)
QY 3 LysLysGlySerValIleValGlyArgIleAsn----- 14
Db 3354 CGTAGGGCCAGAGATAGTCTGGCCAGCGGAGGATGCTCCAGGGGTGGAGG 3413
QY 15 ---LeuSerGlyAspThrAlaThrAlaGlnGlnThrArgGlyGluGluCysGlnGlu 33
Db 3414 TTGCTGGCGCCCATCAGCGGTACGCCAGCCAGCAGAGAGGCGCTCTAGGGGTATATC 3473
QY 34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 53

Db 3474 ACACGCTGACTGGCGGCGACAAACCAAGTGGAGGTGAGGTCCAGTCTGTCAACT 3533
QY 54 AlaThrGlnIlePheLeuAlaThrCysIleAsnGlyValCysIleThrValThrHisGly 73
Db 3534 GCTAGCCAAACCTTCTGCGCAACGTGCATCAATGGGTATGCTGACCTGTACACCGG 3593
QY 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValThrGlnMetThrAsnVal 93
Db 3594 GCGGGAACGAGGACCATCGATCAACCCAGGCTCTGTATCCAGATGATACCAATGTG 3653
QY 94 AspLysAspLeuValGlyTrpGlnAlaProGlnGlySerArgSerLeuThrProCysThr 113
Db 3654 GACCAAGACCTTGTGGGTGGCGCTCTCAAGTTCCCGCTATTGACACCTGCACACC 3713
QY 114 CysGlySerSerAspLeuThrLeuValThrArgHisAlaAspValIleProValArgArg 133
Db 3714 TGGGCTCTCGGACCTTTACCTGTGTACGAGGACGCGGATGCTATCCGTCGCGCGG 3773
QY 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerThrLysGlySer 153
Db 3774 CGAGGTGATAGCAGGGTAGCTGCTTTCGCCCGGCCCATTTCTTCTACTTGAAGGCTCC 3833
QY 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaVal 173
Db 3834 TCGGGGGTCCGCTGTGTGCGCGCGGACACGCGCTGGGCTATTTCAGGCGCGCGTG 3893
QY 174 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThr 193
Db 3894 TGACCCCGTGGAGTGGCTAAGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACACC 3953
QY 194 MetArgSerPro 197
Db 3954 ATGAGATCCCG 3965
RESULT 15
US-09-995-937-5
; Sequence 5, Application US/0995937
; Publication No. US20030028010A1
; GENERAL INFORMATION:
; APPLICANT: RICE, CHARLES et al.
; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAERKAMP, L.C.
; STREET: 7733 FORSYTH BLVD., SUITE 1400
; CITY: ST. LOUIS
; STATE: MO
; COUNTRY: USA
; ZIP: 63105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/995.937
; FILING DATE: 28-May-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/034,756
; FILING DATE: 04-May-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 6029-4831
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 314-727-5188
; TELEFAX: 314-727-6092
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12980 base pairs

```
;
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-995-937-5

Alignment Scores:
Pred. No.: 4.06e-92 Length: 12980
Score: 888.50 Matches: 172
Percent Similarity: 89.22% Conservative: 10
Best Local Similarity: 84.31% Mismatches: 13
Query Match: 86.09% Indels: 9
DB: 11 Gaps: 1

US-09-965-594-16 (1-197) x US-09-995-937-5 (1-12980)

QY 3 LysLysGlySerValValIleValGlyArgIleAsn----- 14
DB 3354 CGTAGGGCCAGGAGATACTGCTTGGCCGACGCGGAATGCTCTCCAAGGGGTGGAGG 3413
QY 15 ---LeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlu 33
DB 3414 TTGCTGGGCCCATCACGGGCTACGCCACGACGACGAGAGGCTCTCTAGGGTGTATAATC 3473
QY 34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 53
DB 3474 ACCAGCTGACTGGCCGGGACAAAACCAAGTGGAGGGTGAGGTCCAGATCGTGTCAACT 3533
QY 54 AlaThrGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGly 73
DB 3534 GCTACCCAAACCTTCCTGGCAACGTGCATCAATGGGGTATGCTGGACTGCTACCAACGG 3593
QY 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValThrGlnMetTyrThrAsnVal 93
DB 3594 GCGGGAACGAGGACCATCGCATCCCAAGGGTCCTGTATCCAGATGTATACCAATGTG 3653
QY 94 AspLysAspLeuValGlyTrpGlnAlaProGlnGlySerArgSerLeuThrProCysThr 113
DB 3654 GACCAAGACCTTGTGGCTGGCCCGCTCCTCAAGGTTCGGCTCATTTGACACCCCTGCACC 3713
QY 114 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 133
DB 3714 TGGGGCTCTCGGACCTTTACCTGGTCACGAGGACGCGCATGTCATCCCGTGGCGCGG 3773
QY 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSertTyrLeuLysGlySer 153
DB 3774 CGAGTGNTAGCAGGGGTAGCCTGCTTTCGCCCGGCCCATTTCTTACTTGAAGGCTCC 3833
QY 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaIleVal 173
DB 3834 TCGGGGGTCCGCTGTGTGTCGCCCGGGACACGCGGTGGCGCTATTTCAGGGCGCGGTG 3893
QY 174 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 193
DB 3894 TGCACCCGTGGAGTGGCTAAGCGGTGGACTTTATCCCTCTGGAGAACCCTAGAGACACC 3953
QY 194 MetArgSerPro 197
DB 3954 ATGAGATCCCCG 3965
```

Search completed: August 31, 2003, 04:54:21
Job time : 190.482 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 30, 2003, 19:20:43 ; Search time 1910.31 Seconds
(without alignments)
2506.388 Million cell updates/sec

Title: US-09-965-594-16
Perfect score: 1032
Sequence: 1 MKKKGSVVIGRINLSGDTA.....VAKAVDFIPVESLETTMRSP 197

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_p2n_model -DEV=xlp
-O=/cpn2_1/USPTO_spool/US09965594/runat_29082003_151919_28322/app_query.fasta_1.2872
-DB=EST -QEXT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09965594 -CGN_1_1_12630_@runat_29082003_151919_28322 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCAPOP=6
-FCAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:**
1: em_estba:**
2: em_esthum:**
3: em_estin:**
4: em_estnu:**
5: em_estov:**
6: em_estpl:**
7: em_estro:**
8: em_htc:**
9: gb_est1:**
10: gb_est2:**
11: gb_hcc:**
12: gb_est3:**
13: gb_est4:**
14: gb_est5:**
15: em_estfun:**
16: em_estom:**
17: em_gss_hum:**
18: em_gss_inv:**
19: em_gss_pln:**
20: em_gss_vrt:**
21: em_gss_fun:**
22: em_gss_man:**
23: em_gss_mus:**
24: em_gss_pro:**
25: em_gss_rod:**
26: em_gss_phg:**
27: em_gss_vrl:**
28: gb_gssl:**

29: gb_gss2:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
c 1	101.5	9.6	1403	13	BQ926101	BQ926101 AGENCOURT
c 2	99.5	9.6	779	10	BF631437	BF631437 HVSMEB001
c 3	98.5	9.5	1199	13	BQ892487	BQ892487 AGENCOURT
c 4	98	9.5	846	10	BF182274	BF182274 G01804028
c 5	98	9.5	984	10	BF304599	BF304599 G01888252
c 6	97.5	9.4	905	13	BQ542842	BQ542842 AGENCOURT
c 7	97	9.4	1141	11	AK080545	AK080545 Mus muscu
c 8	96	9.3	615	12	BQ001625	BQ001625 BJ001625
c 9	96	9.3	643	12	BQ024121	BQ024121 BJ024121
c 10	96	9.3	754	12	BJ016176	BJ016176 BJ016176
c 11	96	9.3	1031	14	CB950999	CB950999 AGENCOURT
c 12	95	9.2	580	14	CA728398	CA728398 wdlc.pk0
c 13	95	9.2	961	10	BF203316	BF203316 G01865914
c 14	95	9.2	1146	12	BM915803	BM915803 AGENCOURT
c 15	94	9.1	701	10	BF863244	BF863244 G63042C02
c 16	94	9.1	1100	10	BG420390	BG420390 G02452419
c 17	94	9.1	1101	29	BZ567280	BZ567280 pacas2-164
c 18	93.5	9.1	898	14	CA787713	CA787713 AGENCOURT
c 19	93.5	9.1	1505	10	BF183416	BF183416 G01809557
c 20	93	9.0	644	29	BX238988	BX238988 Danio rer
c 21	93	9.0	1411	11	BC020343	BC020343 Homo sapi
c 22	92.5	9.0	488	10	BF776637	BF776637 287489 MA
c 23	92.5	9.0	993	9	AL555424	AL555424 AL555424
c 24	92.5	9.0	1329	13	BQ960995	BQ960995 AGENCOURT
c 25	92	8.9	649	10	BE289911	BE289911 G01089126
c 26	92	8.9	757	12	BI258851	BI258851 G02969666
c 27	91.5	8.9	502	9	AA036834	AA036834 zk29405.r
c 28	91.5	8.9	539	10	BE757615	BE757615 212104 MA
c 29	91.5	8.9	844	11	CNS0904S	BNX053096 Single re
c 30	91.5	8.9	844	12	BI198486	BI198486 G02760491
c 31	91.5	8.9	938	13	BQ894657	BQ894657 AGENCOURT
c 32	91.5	8.9	1000	13	BQ735135	BQ735135 AGENCOURT
c 33	91	8.8	470	13	BQ758584	BQ758584 EBma07_SO
c 34	91	8.8	471	13	BU978992	BU978992 HA14N15T
c 35	91	8.8	515	14	CA023748	CA023748 HZ47E17r
c 36	91	8.8	583	12	BM374064	BM374064 HVSMEC001
c 37	91	8.8	938	10	BG309750	BG309750 HVSMEC001
c 38	91	8.8	1283	13	BQ709745	BQ709745 AGENCOURT
c 39	90.5	8.8	528	28	AQ620249	AQ620249 HS_5182_B
c 40	90.5	8.8	577	29	BZ904625	BZ904625 CH240_261
c 41	90.5	8.8	2904	11	AK046602	AK046602 Mus muscu
c 42	90	8.7	460	14	CB883286	CB883286 HQ01M02W
c 43	90	8.7	500	12	BM708007	BM708007 UI-E-C11-
c 44	90	8.7	553	9	AL856435	AL856435 AL856435
c 45	90	8.7	569	12	BM825317	BM825317 R-EST0097

ALIGNMENTS

RESULT 1
LOCUS BQ926101/c
DEFINITION AGENCOURT_8752655 NIH_MGC_130 Mus musculus CDNA clone IMAGE:6335718
5', mRNA sequence.
ACCESSION BQ926101
VERSION BQ926101.1 GI:22341132
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1403)


```

VERSION      BF182274.1  GI:11060416
KEYWORDS     EST.
SOURCE       Mus musculus (house mouse)
ORGANISM
REFERENCE    1 (bases 1 to 846)
AUTHORS      NIH-MGC http://mgi.nci.nih.gov/
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished
COMMENT      Contact: Robert Strausberg, Ph.D.
              Email: cgapbs@mail.nih.gov
              Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: L1AM9308 row: g column: 07
              High quality sequence stop: 696.
              Location/Qualifiers
FEATURES     1..846
              /organism="Mus musculus"
              /mol_type="mRNA"
              /strain="C57BL/6J"
              /db_xref="taxon:10090"
              /clone="IMAGE:4035102"
              /tissue_type="tumor, gross tissue"
              /dev_stage="7 months"
              /lab_host="DH10B"
              /clone_lib="NCI-CGAP_Mam5"
              /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
              site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
              Library constructed by Life Technologies. Investigators
              providing samples: Lothar Hennighausen/Robin Humphreys,
              NIH"
BASE COUNT   176 a 218 c 241 g 210 t 1 others
ORIGIN
Alignment Scores:
Pred. No.:    17.6      Length:      846
Score:        98.00     Matches:    46
Percent Similarity: 48.00%  Conservative: 14
Best Local Similarity: 36.80%  Mismatches: 35
Query Match:  9.50%    Indels:    31
Db:           10       Gaps:      9
US-09-965-594-16 (1-197) x BF182274 (1-846)
QY 73 GlyAlaGlyThrArg-ThrIleAlaSerProLysGlyProValThrGlnMetTyrThrAs 92
DB 757 GGTCTGCTACCAAGACAGCGCTGGATGACAGAAAGGACCA-----CATCCTTC 710
QY 92 nValAspLysAspLeuValGlyTrpGln-----AlaProG1 104
DB 709 GGTCTTCAGTCCCAAGTGGCTGGAGAAAGTGAACACACAGAAAGAGGACGTCCTCA 650
QY 104 nGlySerArg-----SerLeuThrProCysThrCysGlySerSerAspLeuTyrLeuVa 122
DB 649 GTCCCCAGCTTCAGTACTGACACAAAGTGTCTGTGGA-----610
QY 122 lThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeuLeu 142
DB 609 -ACTAGACACACCT--GTAATCCAGGAGGAACCGTGGAGCAACAGAGGACCTCC--CT 556
QY 142 UserProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeu---LeuCysPr 161
DB 555 CACCCCACTCC---TCCGCTCAGCGGCGACCTCTCTCTGGCCGCCACCTCCCTCTGTC 499
QY 161 oAlaGlyHis---AlaValGlyIlePheArg-----AlaAlaValCysThrAr 176
DB 498 TAGTGGGNCCTCTCCCCAGCACCACACAGACTGTACTCCCTTTGGCCCTCTGCACCTCT 439

```

```

QY 176 gGlyValAlaLys 180
DB 438 TGGGATGACTGAG 426
RESULT 5
BF304699/c
LOCUS      601888252f1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:412276 5',
DEFINITION mRNA sequence.
ACCESSION  BF304699
VERSION    BF304699.1  GI:11251586
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 984)
AUTHORS    NIH-MGC http://mgi.nci.nih.gov/
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: L1CM1005 row: g column: 13
            High quality sequence stop: 646.
            Location/Qualifiers
FEATURES     1..984
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="IMAGE:412276"
              /tissue_type="rhabdomyosarcoma"
              /lab_host="DH10B (phage-resistant)"
              /clone_lib="NIH_MGC_17"
              /note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI;
              Site_2: XhoI; cDNA made by oligo-dt priming.
              Directionally cloned into EcoRI/XhoI sites using the
              following 5' adaptor: GGCAGAG(G). Size-selected >500bp
              for average insert size 1.8kb. Library constructed by
              Ling Hong in the laboratory of Gerald M. Rubin (University
              of California, Berkeley) using ZAP-CDNA synthesis kit
              (Stratagene) and Superscript II RT (Life Technologies)."
```

```

BASE COUNT   133 a 329 c 351 g 171 t
ORIGIN
Alignment Scores:
Pred. No.:    21.4      Length:      984
Score:        98.00     Matches:    32
Percent Similarity: 44.05%  Conservative: 5
Best Local Similarity: 38.10%  Mismatches: 25
Query Match:  9.50%    Indels:    22
Db:           10       Gaps:      5
US-09-965-594-16 (1-197) x BF304699 (1-984)
QY 100 TrpGlnAlaProGlnGlySerArgSerLeuThr---ProCysThrCysGlySerSerAsp 118
DB 646 TGGCCAGTCACGGCATTCCTCGTGGAGAGGAGGACCGGTGACCTGC-----599
QY 119 LeuTyrLeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArg 138
DB 598 -----ACCGAGCAGCGAGGACATACATACGAGGAGAGCGTGT---TCCCGC 554
QY 139 GlySerLeuLeuSerProArgPro-----IleSerTyrLeuLysGlySer 153
DB 553 GGGCGCCTCTTGTGGGAGAGACCTCGATGTGTGTCCAGGTCCGGTGTGTACTGGAAGT 494

```

```

QY 154 SerGlyCysProLeuLeuCysProAlaGlyHisAlaValGlyLePheArgAlaVal 173
Db 493 CGCAGCGCTCCGTCAGTGCAGC-----TTCCAGCGCCGGGG 455
QY 174 CysThrArgGly 177
Db 454 TGGCCGGGAGGA 443

RESULT 6
BU542842/c
LOCUS BU542842 905 bp mRNA linear EST 13-SEP-2002
DEFINITION AGENCOURT_10334715 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6574789
S, mRNA sequence.
ACCESSION BU542842
VERSION BU542842.1 GI:22853325
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 905)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
JOURNAL NIH-MGC http://mgs.nci.nih.gov/.
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: DCTD/DTF
cDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM2770 row: k column: 13
High quality sequence stop: 633.
Location/Qualifiers
1..905
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6574789"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_40"
/note="Organ: prostate; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 201 a 260 c 273 g 171 t
ORIGIN

Alignment Scores:
Pred. No.: 21.5 Length: 905
Score: 97.50 Matches: 51
Percent Similarity: 31.40% Conservative: 14
Best Local Similarity: 24.64% Mismatches: 75
Query Match: 9.45% Indels: 67
DB: 13 Gaps: 9

US-09-965-594-16 (1-197) x BU542842 (1-905)
QY 28 GluGluGlyCysGlnGluThrSerGln-----ThrGlyArgAspLysAsnGlnValGluGly 46
Db 884 AAGAGAGGCGCCCGCTGCTGTTCCCGAGGAAGGGGACCCGAGACCAAGAGGAGGAGGCGC 825
QY 47 GluValGlnIleValSerThrAlaThrGlnThrPheLeuAla----- 60
Db 824 GGGGCGCTCTTCCAGGCGCCCTGTGCACAAAGTGTCCTTGGGTGCGCCGCGCCATGTCCTCA 765

```

```

QY 61 -----ThrCysIleAsnGlyValCys----- 67
Db 764 CATTTTCGACATCCGCGCAGAACATGTGTGGGTCTTGTCCCGCAGCAGGCGACGCC 705
QY 68 ---TrpThrValTyrHisGlyAla----- 74
Db 704 AAGTGGGAGGAGGCGCATGTGTGCACACGCTGGGGAGGCGCCCTGGTGAGAAGCAGCCCA 645
QY 75 -----GlyThrArgThrIle 79
Db 644 CAGTAGCAGCCCATCCAGAGGAGACCATCTCCGAGGGCCACAGGCCCTCTGCAGCCCTG 585
QY 80 AlaSerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGly 99
Db 584 GCACGTGCGCGCCAGCCCTCCATCTCAGCGGATGTGCAGGAGTGAGACAGCAATCCAGGA 525
QY 100 TrpGlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeu 119
Db 524 CGTTCTGCCCTAGGTGACGCTCTTCATCCGCCCTGTTGTGCTTGCATGCTCAAGGTTG 465
QY 120 -----TyrLeuValThrArgHisAlaAsp-----ValIleProValArgArg 133
Db 464 CCCGTGTCACAGCTGTGCAACGCCATCCAGGGGCTTCGTTGTCTCTCCAGCTCACTCT 405
QY 134 Arg-----GlyAspSerArgGlySerLeuLeuSerPro-----Arg 145
Db 404 CGGCTCCAGGCGCCAGCCCTTCATCTCCTCAGGATCTGGGTTAGTTCTCTGGGTATCTG 345
QY 146 ProIleSerTyrLeuLysGlySerSerGlyClyProLeuLeuCysProAlaGly----- 163
Db 344 CCTCAGAAAGGGCTGGCAGGCTTGTCTGCAGGTGCAGGTGCTGCTGCCCTCTGCTCTCT 285
QY 164 -----HisAlaValGly 167
Db 284 CGGTGGCTCAGCGTGCCAGGG 264

RESULT 7
AK080545 1141 bp mRNA linear HTC 05-DEC-2002
LOCUS AK080545
DEFINITION Mus musculus 7 days neonate cerebellum cDNA, RIKEN full-length
enriched library, clone:A730082L10 product:weakly similar to zinc
finger protein (fragment) [Mus musculus], full insert sequence.
ACCESSION AK080545.1 GI:26348600
VERSION AK080545
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Alizawa,K., Nagaoaka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer

```

```

JOURNAL      Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE      20530913
PUBMED      11076861
REFERENCE    4
AUTHORS      Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
              Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
              Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaoka,I.,
              Saito,T., Okazaki,Y., Goshiori,T., Bono,H., Kasukawa,T., Saito,R.,
              Kadota,H., Matsuda,H., Ashtburner,M., Batalov,S., Casavant,T.,
              Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
              Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
              Quackenbush,J., Schriml,L.M., Staabli,F., Suzuki,R., Tomita,M.,
              Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
              Balarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
              Carninci,P., de Bernaldo,M.F., Brownstein,M.J., Bult,C.,
              Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
              Hofmann,M., Hume,D.A., Kaniya,M., Lee,N.H., Lyons,P.,
              Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
              Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
              Sato,K., Schonbach,C., Seva,T., Shibata,Y., Storch,K.F., Suzuki,H.,
              Toyooka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
              Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.,
              and Hayashizaki,Y.
TITLE        A full-length mouse cDNA collection
JOURNAL      Nature 409 (6821), 685-690 (2001)
MEDLINE      21085660
PUBMED      11217851
REFERENCE    5
AUTHORS      The FANTOM Consortium and the RIKEN Genome Exploration Research
              Group Phase I & II Team.
TITLE        Analysis of the mouse transcriptome based on functional annotation
JOURNAL      Nature 420, 563-573 (2002)
MEDLINE      12117851
PUBMED      12117851
REFERENCE    6
AUTHORS      Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
              Fukuda,S., Furuno,M., Hamae,K., Hiramoto,K., Hirose,T., Hirozane,T.,
              Hayashida,K., Hayatsu,N., Kondo,S., Kondo,S., Konno,H., Kouda,M.,
              Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
              Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
              Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
              Okazaki,Y., Saito,R., Shibata,K., Shinagawa,A., Shiraki,T.,
              Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
              Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akashira,S.,
              Tanaka,Y., Tanaka,T., Tomaru,A., Toyota,T., Yasunishi,A.,
              Muramatsu,M. and Hayashizaki,Y.
TITLE        Direct Submission
JOURNAL      Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
              Physical and Chemical Research (RIKEN), Laboratory for Genome
              Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
              RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
              Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp,
              URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
              Fax:81-45-503-9216)
COMMENT      cDNA library was prepared and sequenced in Mouse Genome
              Encyclopedia Project of Genome Exploration Research Group in Riken
              Genomic Sciences Center and Genome Science Laboratory in RIKEN.
              Division of Experimental Animal Research in Riken contributed to
              prepare mouse tissues.
              Please visit our web site for further details.
              URL:http://genome.gsc.riken.go.jp/
              URL:http://fantom.gsc.riken.go.jp/.
FEATURES     source
              Location/Qualifiers
                1..1141
                /organism="Mus musculus"
                /mol_type="mRNA"
                /strain="C57BL/6J"
                /db_xref="FANTOM DB:A730082L10"
                /db_xref="taxon:10090"
                /clone="A730082L10"
                /tissue_type="cerebellum"
                /clone_lib="RIKEN full-length enriched mouse cDNA library"
                /dev_stage="7 days neonate"

```

```

<1..587
/note="unnamed protein product; putative
weakly similar to zinc finger protein (fragment) [Mus
musculus] (PIR1148722, evidence: FASTA, 50.7%ID,
57.6%length, match=601)"
/codon_start=3
/protein_id="BAC37940.1"
/db_xref="GI:26348601"
/translation="DSCLPAAAPGSRLLTPRGDFFLKEKLSARAVGPGPSVAFVS
TRVRAQAQCGRRRGRRSEGLSKRPFRHVPVPGVHVLGGRRIPPPAGE
AQAAGRAQQVHPHPGPHGTVPVPOGAAGLLPALAAQVPGVPGREGPRAPRHS
PKPVPTALGFSFGQGGPAPILLAPANGRSVGLAL"
polyA_signal 1118..1123
/note="putative"
1141
polyA_site 244 a 316 c 353 g 228 t
BASE COUNT 244 a 316 c 353 g 228 t
ORIGIN
Alignment Scores:
Pred. No.: 32.5 Length: 1141
Score: 97.00 Matches: 46
Percent Similarity: 39.31% Conservative: 11
Best Local Similarity: 31.72% Mismatches: 53
Query Match: 9.40% Indels: 35
DB: 11 Gaps: 8
US-09-965-594-16 (1-197) x AK080545 (1-1141)
QY 68 TrpThrValTyrHisGlyAlaGly-----ThrArgThrIleala 80
Db 278 TGGCAGACGACATCCCGCGCGCGGAGAGCGGAGCGGCGCGCGCGAGCA 337
QY 81 Ser-ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTr 100
Db 338 AGTGGCGCATCCCGCTGGCGGCGCACATGGAACAGTGTGTG----- 377
QY 100 pGlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTy 120
Db 378 ---CCTCCTCAAGGAGCGGCTGGCTTCTCCTCTCA----- 410
QY 120 rLeuValThrArgHisAlaAspValIleProValArg---ArgArgGlyAspSerArgGl 139
Db 411 -CTCGCAGCTCGCCCAAGTCTCTGTGGCGGTTAGGCGCGGAGGACCAAGAGGCGC 469
QY 139 ySerLeuLeuSerProArgProile-----SertyrLeuLysGlyse 153
Db 470 ACCGAGACACACGCCCAAGCGGTTCTACAGCCTTGGGTTTCTGTTGGCAGGTTGG 529
QY 153 rSerGlyGlyProLeuLeuLysProAla---GlyHisAlaValGlyIlePheArgAlaAl 172
Db 530 GCCTGTCTCTCCTCCTAGCCCCAGCAAGTGGAGGCTGTAGGGTTG---GCCCTGTA 586
QY 172 aValCysThrArgGlyValAlaLysAlaVal-----AspPheIleProValGluSerLe 190
Db 587 AGTGTGTGAGTCGGAGACTTGAGGGTGTGGCTGTGAGTGTGACACCAAGCTCTCTCT 646
QY 190 uGluThrThrMet 194
Db 647 GCGAGGTACACTT 659
RESULT 8
BJ001625/c 615 bp mRNA linear EST 05-DEC-2001
LOCUS BJ001625 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA025C02 5',
DEFINITION mRNA sequence.
ACCESSION BJ001625
VERSION BJ001625.1 GI:17364516
KEYWORDS EST.
SOURCE Oryzias latipes (Japanese medaka)
ORGANISM Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

```


Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.

REFERENCE
1 (bases 1 to 615)
AUTHORS Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
TITLE Medaka EST Project in Takeda's lab
JOURNAL Unpublished

COMMENT
Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES
source
1..615
/organism="Oryzias latipes"
/mol_type="mRNA"
/strain="Hd-rR"
/db_xref="taxon:8090"
/clone="MF01SSA025C02"
/sex="mixture of female and male"
/tissue_type="whole embryo"
/dev_stage="segmentation stage 20 - 25"
/clone_lib="MF01SSA cDNA"
BASE COUNT 140 a 166 c 165 g 144 t
ORIGIN

Alignment Scores:
Pred. NO.: 18.3 Length: 615
Score: 96.00 Matches: 43
Percent Similarity: 34.72% Conservative: 7
Best Local Similarity: 29.86% Mismatches: 58
Query Match: 9.30% Indels: 36
DB: 12 Gaps: 6

US-09-965-594-16 (1-197) x BJ001625 (1-615)

QY 41 LysAsnGlnValGluGlyGluValGlnLeValSerThrAlaThrGlnThrPheLeuAla 60
||||| ||||| :|||
DB 511 AAAATGACGTAGTACCAAGACACACATCCACACACATGTTCTGTTCTACGGGCT 452
QY 61 ThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
||||| ||||| :|||
DB 451 -----TGTGGAGAACCTATCATCTCTGCTTTAGAGCAACGGCA 410
QY 81 SerProLys-----GlyProValThrGlnMetTyrThrAsnValAspLys 95
||| :||| |||||
DB 409 GCTCTGGCGGCGAGGAGCTCTGGCCAGTGTGACTCGT----- 368
QY 96 AspLeuValGlyTrpGlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGly 115
||| :||| |||||
DB 367 -----GGAGGACGAAGACGCTCACCCGGAGCTGTAGGCTGCAGGATGCGGATGTGGC 314
QY 116 SerSerAspLeuTyrLeuValThrArg----- 124
||||| |||||
DB 313 TCTGCT-----TTGGTCTCTGCTCTCTGATCATCTTCTCATCTGACCTTCCA 263
QY 125 HisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeuLeuSerPro 144
||| ||| |||||
DB 262 CATCCAGGTGTGCCAGCGCTGCTGCGGGTGTGGAGAGCGCGACAGGCAGCAGT 203
QY 145 Arg-----ProIleSerTyrLeuLysGlySerGlyGlyProLeuLeuCysPro 161
||| ||| |||||
DB 202 CGGGGGTGAATCTCTGCAGGAGCTTTCACGCGGATCATGAGGAGCAGCTCGCTGCACA 143
QY 162 AlaGlyHisAla 165
|||
DB 142 GCCTCTGTGTGCA 131

RESULT 9
BJ024121
LOCUS BJ024121 MF01SSA cDNA Oryzias latipes cDNA linear EST 05-DEC-2001
DEFINITION BJ024121 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA143D12 3',

mRNA sequence.
BJ024121
BJ024121.1 GI:17377389
EST.
Oryzias latipes (Japanese medaka)
Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
1 (bases 1 to 643)
Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
TITLE Medaka EST Project in Takeda's lab
JOURNAL Unpublished
COMMENT
Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES
source
1..643
/organism="Oryzias latipes"
/mol_type="mRNA"
/strain="Hd-rR"
/db_xref="taxon:8090"
/clone="MF01SSA143D12"
/sex="mixture of female and male"
/tissue_type="whole embryo"
/dev_stage="segmentation stage 20 - 25"
/clone_lib="MF01SSA cDNA"
BASE COUNT 171 a 148 c 148 g 176 t
ORIGIN

Alignment Scores:
Pred. NO.: 19.4 Length: 643
Score: 96.00 Matches: 43
Percent Similarity: 34.72% Conservative: 7
Best Local Similarity: 29.86% Mismatches: 58
Query Match: 9.30% Indels: 36
DB: 12 Gaps: 6

US-09-965-594-16 (1-197) x BJ024121 (1-643)

QY 41 LysAsnGlnValGluGlyGluValGlnLeValSerThrAlaThrGlnThrPheLeuAla 60
||||| ||||| :|||
DB 242 AAAATGACGTAGTACCAAGACACACATCCACACACATGTTCTGTTCTACGGGCT 301
QY 61 ThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
||||| ||||| :|||
DB 302 -----TGTGGAGAACCTATCATCTCTGCTTTAGAGCAACGGCA 343
QY 81 SerProLys-----GlyProValThrGlnMetTyrThrAsnValAspLys 95
||| :||| |||||
DB 344 GCTCTGGCGGCGAGGAGCTCTGGCCAGTGTGACTCGT----- 385
QY 96 AspLeuValGlyTrpGlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGly 115
||| :||| |||||
DB 386 -----GGAGGACGAAGACGCTCACCCGGAGCTGTAGGCTGCAGGATGCGGATGTGGC 439
QY 116 SerSerAspLeuTyrLeuValThrArg----- 124
||||| |||||
DB 440 TCTGCT-----TTGGTCTCTGCTCTCTGATCATCTTCTCATCTGACCTTCCA 490
QY 125 HisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeuLeuSerPro 144
||| ||| |||||
DB 491 CATCCAGGTGTGCCAGCGCTGCTGACGCGTGTGGAGAGCGCGACAGGCAGCAGT 550
QY 145 Arg-----ProIleSerTyrLeuLysGlySerGlyGlyProLeuLeuCysPro 161
||| ||| |||||
DB 551 CGGGGGTGAATCTCTGCAGGAGCTTTCACGCGGATCATGAGGAGCAGCTCGCTGCACA 610

QY	60	AlaThrCysIleAsnGlyValCysTrpThrVal-----TyrHisGlyAlaGlyThrArg	77
Ddb	40 <td>GCATGGTGTTCTTCTGCTTCCTGGTGGACCGCGGAAGGTGCGGAGCAGCAGCGG</td> <td>99</td>	GCATGGTGTTCTTCTGCTTCCTGGTGGACCGCGGAAGGTGCGGAGCAGCAGCGG	99
QY	78 <td>ThrIleAlaSerProIysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeu</td> <td>97</td>	ThrIleAlaSerProIysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeu	97
Ddb	100 <td>CGCGGGGATCTGCTCGCGTGGCGGCTGCGCAGGCTGCGGACA-----</td> <td>147</td>	CGCGGGGATCTGCTCGCGTGGCGGCTGCGCAGGCTGCGGACA-----	147
QY	98 <td>ValGlyTrpGlnAlaProGlnGlySerArgSerLeuThrProCysThrCys-----Gly</td> <td>115</td>	ValGlyTrpGlnAlaProGlnGlySerArgSerLeuThrProCysThrCys-----Gly	115
Ddb	148	-----TGGAGAGCGGCACCGGCTCTGCTACTCTCTCACCCTGACCGCGCTCACCTGGC	201
QY	116 <td>SerSerAsp-LeuTyrIleuValThrArgHisAlaAspValIleProVal-----</td> <td>131</td>	SerSerAsp-LeuTyrIleuValThrArgHisAlaAspValIleProVal-----	131
Ddb	202 <td>CGCGCATCATCTGCACCTTCTGGGGGCGCATCTCAAGTCTCCTACCGCCTACAGGCTCT</td> <td>261</td>	CGCGCATCATCTGCACCTTCTGGGGGCGCATCTCAAGTCTCCTACCGCCTACAGGCTCT	261
QY	132	-----ArgArgArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSer	148
Ddb	262 <td>ACTAGACGGCGGATTGATCTGCGCGCGGCGAGCAGGTCTCCACCGCGCTCCCGTTCC</td> <td>319</td>	ACTAGACGGCGGATTGATCTGCGCGCGGCGAGCAGGTCTCCACCGCGCTCCCGTTCC	319

RESULT	13	961 bp	mRNA	linear	EST 06-NOV-2000
LOCUS	BF203316/c	601865914F1	NIH_MGC_17	Homo sapiens	cDNA clone IMAGE:4098578 '5',
DEFINITION	mRNA sequence.				
ACCESSION	BF203316				
VERSION	1				
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 961)				
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC) unpublished				
JOURNAL	Contact: Robert Strausberg, Ph.D.				
COMMENT	Email: cgapbs-r@mail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Ling Hong/Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov Plate: L1CM965 row: 1 column: 03 High quality sequence stop: 637.				
FEATURES	<p>1. 961</p> <p>Location/Qualifiers</p> <p>/organism="Homo sapiens"</p> <p>/mol_type="mRNA"</p> <p>/db_xref="taxon:9606"</p> <p>/clone="IMAGE:4098578"</p> <p>/tissue_type="rhadomyosarcoma"</p> <p>/lab_host="DH10B (phage-resistant)"</p> <p>/clone_lib="NIH_MGC_17"</p> <p>/note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI; Site_2: XhoI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."</p>				
BASE COUNT	230 a	297 c	300 g	134 t	
ORIGIN					
Alignment Scores:	40.8	Length:	961		
Pred. No.:	95.00	Matches:	28		
Score:	55.07%	Conservative:	10		
Percent Similarity:					

Best Local Similarity: 40.58% Mismatches: 27
 Query Match: 9.21% Indels: 4
 DB: 10 Gaps: 1

US-09-965-594-16 (1-197) x BF203316 (1-961)

QY 114 CysGlySerSer-AspLeuThrValThrArgHisAlaAspValIleProValArgAR 133
 DB 612 TGTGGAGCTCAGACATCGTCTAGTCGAGGAGCGAGATCGGATCGGAA 553
 QY 133 gArgGlyAspSerArgGlySerLeuLeuSer-ProArgProIleSerTyrLeuLysGlyS 153
 DB 552 GCGCGGGATGCTGATCGGCGATGCTCTCTCGTCCGCGGGGAGTAGCAAGGT 493
 QY 153 eSerGlyCysProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaLav 173
 DB 492 CTGGAGGAGGAGG-----TGCTCGAGTGCTGTAGGTGCTGCCGCTCAGCTT 439
 QY 173 aLCysThrArgGlyValAlaLys 180
 DB 438 TTTGTGATGGTGCACGCTGAA 416

RESULT 14

BM915803/3
 LOCUS 1146 bp mRNA linear EST 12-MAR-2002
 DEFINITION AGENCOURT_6639455 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5482056
 5', mRNA sequence.

ACCESSION BM915803.1 GI:19366182

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1146)
 NIH-MGC <http://mgi.nci.nih.gov/>
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: DCTD/DPF
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLCM2007 row: i column: 01
 High quality sequence start: 6
 High quality sequence stop: 256.
 Location/Qualifiers

1..1146

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:5482056"

/tissue_type="amelanotic melanoma, cell line"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH_MGC_41"

/note="Organ: skin; Vector: pOTB7; Site:1: XhoI; Site:2:

EcoRI; cDNA made by oligo-dT priming. Directionally cloned

into EcoRI/XhoI sites using the following 5' adaptor:

GCACGAG(G). Library constructed by Ling Hong in the

laboratory of Gerald M. Rubin (University of California,

Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and

Superscript II RT (Life Technologies). Note: this is a

NIH_MGC Library."

BASE COUNT 169 a 492 c 344 g 141 t

ORIGIN

Alignment Scores: 51.3 Length: 1146

Pred. No.: 95.00 Matches: 45

Percent Similarity: 34.62% Conservative: 18
 Best Local Similarity: 24.73% Mismatches: 62
 Query Match: 9.21% Indels: 57
 DB: 12 Gaps: 10

US-09-965-594-16 (1-197) x BM915803 (1-1146)

QY 22 AlaGlnGlnThrArgGlyGluGluCysGlnGlnThrSer-GlnThrGlyArgAspLy 41
 DB 1098 GCGCAGAGCGGTGTCGCGAGCGAGGGTCCCTCCGCTCTCTACCTCCGTCGGAG 1039
 QY 41 sAsnGlnValGluGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaTh 61
 DB 1038 CTGAGGAGCAGAGGT-----CTACGCGCGTGGGTAGGGA 1003
 QY 61 rCysIleAsnGlyValCysTyrThrValTyrHis-----GlyAlaGlyThr-- 76
 DB 1002 CCCCCTGTCGGATGTTGTGG-----TATCACTTCCCGCGCGGGGAGGTACGTG 949
 QY 77 -----ArgThrIleAlaSerPro-----LysGlyProValThrGlnMetTy 90
 DB 948 AGCGAGGGCGCGCGTCCGGCGCGCGCGGGGCGCG----- 903
 QY 90 rThrAsnValAspLysAspLeuValGlyTyrPglAlaProGln-----GlySerAr 107
 DB 902 -----CAGATGTGCGGTGGGAAGCGCCGCTCGCGCGGTGGGGCCAG 859
 QY 107 gSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeuValThrArgHisAlaAs 127
 DB 858 ACTTGCTGTGCTGTTCTCTGGG----- 834
 QY 127 pValIleProValArgArgArgGlyAspSerArgGlySerLeuLeuSerProArgProIl 147
 DB 833 -----CGGAGGGCGCGCGGTAGTGTGTGGCGCGGTCCCTCT 787
 QY 147 eSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGl 167
 DB 786 CGGTATCTACAGGCGCGCGGACGACCATCTCTCTCCG-----TG 742
 QY 167 yIlePheArgAlaAlaValCysThrArgGlyValAlaLysAlaValAspPhe---IlePr 186
 DB 741 GGCCTTCGCGCTGCTGTCTTCGCGGTCTGCCGCGGGGGGGGGTTCGCGTACC 682
 QY 186 oval 187
 DB 681 TTG 678

RESULT 15

BF863244

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BF863244 701 bp mRNA linear EST 19-JAN-2001
 963042C02.x1 C. reinhardtii CC-1690, Stress condition I, normalized
 , Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.

BF863244

BF863244.1 GI:12253388

EST.

Chlamydomonas reinhardtii

Chlamydomonas reinhardtii

Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;

Chlamydomonadales; Chlamydomonas.

1 (bases 1 to 701)

Grossman, A., Davies, J., Federspiel, N., Harris, E., Hauser, C.,

Lefebvre, P., McDermott, J. P., Shrager, J., Silflow, C. and Stern, D.

Analyses of the Chlamydomonas reinhardtii Genome: A Model,

Unicellular System for Analyzing Gene Function and Regulation in

Vascular Plants; project phase 3

Unpublished

Contact: Charles Hauser

DCMB Box 91000

Duke University

Durham, NC 27708-1000

Tel: 919 613 8159

Fax: 919 613 8177

Email: chauser@duke.edu.

FEATURES

SOURCE

Location/Qualifiers
 1. .701
 /organism="Chlamydomonas reinhardtii"
 /mol_type="mRNA"
 /strain="CC-1690 wild type mt+ 2lgr"
 /db_xref="taxon:3055"
 /clone_lib="C. reinhardtii CC-1690, Stress condition I,
 normalized, Lambda Zap II"
 /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
 XhoI; This library, constructed by John Davies and Jeffrey
 McDermott, combines cDNAs from CC-1690 cells grown to
 mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min,
 1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr
 , 4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was
 purified from each sample, pooled and cDNA synthesized.
 The cDNA was directionally cloned into lambda zap II
 (Stratagene) in the EcoRI (5') and XhoI (3') sites.
 pBluescript II SK- plasmids were excised from the lambda
 Zap clones by superinfection with ExAssist (Stratagene)
 phage. The library was normalized using method 4 described
 in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT 173 a 213 c 175 g 140 t
 ORIGIN

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
33.9	94.00	32	32
Percent Similarity:	40.71%	Conservative:	14
Best Local Similarity:	28.32%	Mismatches:	45
Query Match:	9.11%	Indels:	22
DB:	10	Gaps:	4

US-09-965-594-16 (1-197) x BF863244 (1-701)

Qy	71	TyrHisGlyAlaGlyThrArgThrIleAlaSerProLys-----GlyProVal	86
Db	171	CAACACCATACCCCTTGCCTCAGCTGCTCACCACCAAAATTATGCCCATACGGGCCACTA	230
Qy	87	ThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGlnAlaProGlnGlySer	106
Db	231	ACAAAGTTACATACACGG-----AAGGACCGCGCGCTTGGCCACCCCTTGGAGCCG	284
Qy	107	ArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeuValThrArgHisAla	126
Db	285	AGAAGCCGACCGTCTCTCTGGGTCAICCCCATGCTATCAATCTCCGCTATCAG	344
Qy	127	AspValIle-----ProValArgArgGlyAspSerArg-----	138
Db	345	GAGATCAITTTGCATGTGGCTTTAGTACCCCAAGAGCGCTGGGAGTGGGCATTTATAA	404
Qy	139	-----GlySerLeuLeuSerProArgProIleSerTyrLeu	150
Db	405	GAAGGGACGGGAATTCGGTTTGGGAAAGTCAGCGCCCAAGGCTGACCAAGTGCTA	464
Qy	151	LysGlySerSerGlyGlyProLeuLeuCysProAlaGly	163
Db	465	CTCCAGGCAGCAATGGGAGCCTTTCGCGGTGTCGGGT	503

Search completed: August 31, 2003, 04:27:34
 Job time : 1916.31 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 17:42:58 ; Search time 44.6227 Seconds
(without alignments)
700.745 Million cell updates/sec

Title: US-09-965-594-18

Perfect score: 1017

Sequence: 1 MKKGSVIVGRINLSGDTA.....VAKAVDHPVESLETMRSP 197

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

```

1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1017	100.0	197	21	AA15223
2	1005	98.8	197	21	AA15224
3	1002	98.5	197	21	AA15222
4	985	97.8	197	21	AA15225
5	985	96.9	197	21	AA15221
6	951	93.5	195	21	AA15220
7	946	93.0	197	21	AA15226
8	912	89.7	195	21	AA15212
9	881.5	86.7	665	20	AA124943
					HCV NS4A-NS3 compl

10	878.5	86.4	665	20	AA124947	HCV NS4A-NS3 compl
11	877.5	86.3	665	20	AA124941	HCV NS4A-NS3 compl
12	877.5	86.3	665	20	AA124942	HCV NS4A-NS3 compl
13	874.5	86.0	216	20	AA117880	HCV NS4A-NS3 compl
14	874.5	86.0	665	20	AA124945	HCV NS4A-NS3 compl
15	874.5	86.0	665	20	AA124946	HCV NS4A-NS3 compl
16	873.5	85.9	665	20	AA124940	HCV NS4A-NS3 compl
17	873.5	85.9	671	20	AA124948	HCV NS4A-NS3 compl
18	871.5	85.7	216	20	AA117884	HCV NS4A-NS3 compl
19	870.5	85.6	216	20	AA117879	HCV NS4A-NS3 compl
20	870.5	85.6	216	20	AA117878	HCV NS4A-NS3 compl
21	870.5	85.6	665	20	AA124944	HCV NS4A-NS3 compl
22	870.5	85.6	671	20	AA124949	HCV NS4A-NS3 compl
23	870	85.5	215	20	AA117890	HCV NS4A-NS3 compl
24	867.5	85.3	216	20	AA117882	HCV NS4A-NS3 compl
25	867.5	85.3	216	20	AA117883	HCV NS4A-NS3 compl
26	867.5	85.3	216	20	AA117886	HCV NS4A-NS3 compl
27	866.5	85.2	216	20	AA117877	HCV NS4A-NS3 compl
28	864	85.0	215	20	AA117887	HCV NS4A-NS3 compl
29	863.5	84.9	216	20	AA117881	HCV NS4A-NS3 compl
30	863.5	84.9	216	20	AA117885	HCV NS4A-NS3 compl
31	859	84.5	213	20	AA117888	HCV NS4A-NS3 compl
32	859	84.5	631	20	AAW3482	HCV NS3 protein.
33	858.5	84.4	191	21	AAW44728	Hepatitis C virus
34	858.5	84.4	3011	19	AAW77397	Hepatitis C virus
35	858.5	84.4	3011	24	ABP71460	Amino acid sequenc
36	858.5	84.4	3012	23	AAU99289	Hepatitis C virus
37	855.5	84.1	3011	14	AAW40120	HCV genomic amino
38	854.5	84.0	687	16	AAW79223	pHCV150-encoded se
39	854.5	84.0	1648	16	AAW79221	pHCV176-encoded se
40	854.5	84.0	1766	10	AAW92041	Sequence encoded i
41	854.5	84.0	1786	10	AAW92058	Protein sequence o
42	854.5	84.0	2261	10	AAW90164	Peptide encoded by
43	854.5	84.0	2301	10	AAW92047	Sequence encoded i
44	854.5	84.0	2436	10	AAW92050	Sequence encoded i
45	854.5	84.0	2436	10	AAW90288	Peptide encoded by

ALIGNMENTS

```

RESULT 1
AA15223
ID  AA15223 standard; protein; 197 AA.
XX
AC  AA15223;
XX
DT  19-DEC-2000 (first entry)
XX
DE  Hepatitis C virus NS4A-NS3 fusion protease #5.
XX
KW  Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW  liver failure; liver cancer; mutant; mutain.
XX
OS  Hepatitis C virus.
OS  Synthetic.
XX
PN  W0200040707-A1.
XX
PD  13-JUL-2000.
XX
PF  06-JAN-2000; 2000WO-US00345.
XX
PR  08-JAN-1999; 98US-0115271.
XX
(PRIM ) BRISTOL-MYERS SQUIBB CO.
XX
PI  Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX
WPI: 2000-465976/40.
XX
N-PSDB: AAA73332.
XX
PT  Modified hepatitis C virus (HCV) NS3 protease comprising at least 1

```

PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 PT -
 XX
 XX Claim 23; Fig 15; 66pp; English.
 XX
 CC The present sequence is a mutated version of a fusion protein created
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
 CC proteins are both essential for the replication of the virus, acting to
 CC cleave its replicative proteins from the polyprotein produced from the
 CC HCV genome. Inhibitors of the two proteins should be effective as
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A
 CC fusion proteins which can be used to identify inhibitors of this type, as
 CC well as enabling structural studies of the protease and
 CC protease:inhibitor complexes. This sequence contains the alpha-helix0-1
 CC variant.
 XX
 XX Sequence 197 AA;

Query Match 100.0%; Score 1017; DB 21; Length 197;
 Best Local Similarity 100.0%; Pred. No. 1.6e-96;
 Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 MKKGSVIVGRINLSGDTAYAOOTRGECCQTSOTGRDKNOVEGEVQIVSTATQTFLA 60
 Db 1 MKKGSVIVGRINLSGDTAYAOOTRGECCQTSOTGRDKNOVEGEVQIVSTATQTFLA 60
 Qy 61 TSINGVLWTVYHGAGTETIASPKGPVTOMTYNDKDLVGWQAPQGSRLTPTCGSSDLY 120
 Db 61 TSINGVLWTVYHGAGTETIASPKGPVTOMTYNDKDLVGWQAPQGSRLTPTCGSSDLY 120
 Qy 121 LVTRHADVIPVRRGDSRGSLLSPRISYILKSGSGGPLLCPCAGHAYGIFRAAVSTRGVAK 180
 Db 121 LVTRHADVIPVRRGDSRGSLLSPRISYILKSGSGGPLLCPCAGHAYGIFRAAVSTRGVAK 180
 Qy 181 AVDFIPVESLETTMRSP 197
 Db 181 AVDFIPVESLETTMRSP 197

RESULT 2
 AAB15224
 ID AAB15224 standard; protein: 197 AA.

XX AAB15224;
 XX
 XX 19-DEC-2000 (first entry)
 XX
 XX Hepatitis C virus NS4A-NS3 fusion protease #6.
 XX
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
 KW liver failure; liver cancer; mutant; mutein.
 XX
 OS Hepatitis C virus.
 OS Synthetic.
 XX
 XX WO200040707-A1.
 PN
 PD 13-JUL-2000.
 XX
 PF 06-JAN-2000; 2000WO-US00345.
 XX
 PR 08-JAN-1999; 99US-0115271.
 XX
 XX (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX
 PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
 XX
 XX WPI: 2000-465976/40.
 DR N-PSDB; AAA73333.
 XX

PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 PT -
 XX
 XX Claim 23; Fig 16; 66pp; English.
 XX
 CC The present sequence is a mutated version of a fusion protein created
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
 CC proteins are both essential for the replication of the virus, acting to
 CC cleave its replicative proteins from the polyprotein produced from the
 CC HCV genome. Inhibitors of the two proteins should be effective as
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A
 CC fusion proteins which can be used to identify inhibitors of this type, as
 CC well as enabling structural studies of the protease and
 CC protease:inhibitor complexes. This sequence contains the alpha-helix0-7
 CC variant.
 XX
 XX Sequence 197 AA;

Query Match 98.8%; Score 1005; DB 21; Length 197;
 Best Local Similarity 98.5%; Pred. No. 2.7e-95;
 Matches 194; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 MKKGSVIVGRINLSGDTAYAOOTRGECCQTSOTGRDKNOVEGEVQIVSTATQTFLA 60
 Db 1 MKKGSVIVGRINLSGDTAYAOOTRGECCQTSOTGRDKNOVEGEVQIVSTATQTFLA 60
 Qy 61 TSINGVLWTVYHGAGTETIASPKGPVTOMTYNDKDLVGWQAPQGSRLTPTCGSSDLY 120
 Db 61 TSINGVLWTVYHGAGTETIASPKGPVTOMTYNDKDLVGWQAPQGSRLTPTCGSSDLY 120
 Qy 121 LVTRHADVIPVRRGDSRGSLLSPRISYILKSGSGGPLLCPCAGHAYGIFRAAVSTRGVAK 180
 Db 121 LVTRHADVIPVRRGDSRGSLLSPRISYILKSGSGGPLLCPCAGHAYGIFRAAVSTRGVAK 180
 Qy 181 AVDFIPVESLETTMRSP 197
 Db 181 AVDFIPVESLETTMRSP 197

RESULT 3
 AAB15222
 ID AAB15222 standard; protein: 197 AA.

XX AAB15222;
 XX
 XX 19-DEC-2000 (first entry)
 XX
 XX Hepatitis C virus NS4A-NS3 fusion protease #4.
 XX
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
 KW liver failure; liver cancer; mutant; mutein.
 XX
 OS Hepatitis C virus.
 OS Synthetic.
 XX
 XX WO200040707-A1.
 PN
 PD 13-JUL-2000.
 XX
 PF 06-JAN-2000; 2000WO-US00345.
 XX
 PR 08-JAN-1999; 99US-0115271.
 XX
 XX (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX
 PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
 XX
 XX WPI: 2000-465976/40.
 DR N-PSDB; AAA73331.
 DR

XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 PT
 XX
 PS Claim 23; Fig 14; 66pp; English.
 CC The present sequence is a mutated version of a fusion protein created
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
 CC proteins are both essential for the replication of the virus, acting to
 CC cleave its replicative proteins from the polyprotein produced from the
 CC HCV genome. Inhibitors of the two proteins should be effective as
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A
 CC fusion proteins which can be used to identify inhibitors of this type, as
 CC well as enabling structural studies of the protease and
 CC protease-inhibitor complexes. This sequence contains the alpha-helix0-1
 CC variant.
 XX
 SQ Sequence 197 AA;

Query Match 98.5%; Score 1002; DB 21; Length 197;
 Best Local Similarity 98.5%; Pred. No. 5.6e-95;
 Matches 194; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MKKKGSVIVGRINLSGDTAYAAQOTRGECCQETSQTGRDKNQVEGEVQIVSTATQTFLA 60
 DB 1 MKKKGSVIVGRINLSGDTAYAAQOTRGECCQETSQTGRDKNQVEGEVQIVSTATQTFLA 60
 QY 61 TSINGVLWTVYHGAGTRTIAAPKGPVTQMTYTNVDKDLVGVQAPQGSRLTPTCTCGSSDLY 120
 DB 61 TCINGVCMVTVYHGAGTRTIAAPKGPVTQMTYTNVDKDLVGVQAPQGSRLTPTCTCGSSDLY 120
 QY 121 LVTRHADVIPVRRGDSRGSLLSPRISYLVKSGSGGPLLCAGHAGVGFRAAVSTRGVAK 180
 DB 121 LVTRHADVIPVRRGDSRGSLLSPRISYLVKSGSGGPLLCAGHAGVGFRAAVSTRGVAK 180
 QY 181 AVDFIPVESLETTMRSP 197
 DB 181 AVDFIPVESLETTMRSP 197

RESULT 4
 AAB15225
 ID AAB15225 standard; protein: 197 AA.
 AC AAB15225;
 DT 19-DEC-2000 (first entry)
 DE Hepatitis C virus NS4A-NS3 fusion protease #7.
 DE
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
 KW liver failure; liver cancer; mutant; mutain.
 KW
 OS Hepatitis C virus.
 OS Synthetic.
 OS
 PN WO2000040707-A1.
 XX
 PD 13-JUL-2000.
 XX
 PF 06-JAN-2000; 2000WO-US00345.
 XX
 PR 08-JAN-1999; 99US-0115271.
 XX
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX
 PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
 XX WPI: 2000-465976/40.
 DR

DR N-PSDB: AAA73334.
 XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 PT
 XX
 PS Claim 23; Fig 17; 66pp; English.
 CC The present sequence is a mutated version of a fusion protein created
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
 CC proteins are both essential for the replication of the virus, acting to
 CC cleave its replicative proteins from the polyprotein produced from the
 CC HCV genome. Inhibitors of the two proteins should be effective as
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A
 CC fusion proteins which can be used to identify inhibitors of this type, as
 CC well as enabling structural studies of the protease and
 CC protease-inhibitor complexes. This sequence contains the alpha-helix0-7
 CC variant.
 XX
 SQ Sequence 197 AA;

Query Match 97.8%; Score 995; DB 21; Length 197;
 Best Local Similarity 98.0%; Pred. No. 2.9e-94;
 Matches 193; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MKKKGSVIVGRINLSGDTAYAAQOTRGECCQETSQTGRDKNQVEGEVQIVSTATQTFLA 60
 DB 1 MKKKGSVIVGRINLSGDTAYAAQOTRGECCQETSQTGRDKNQVEGEVQIVSTATQTFLA 60
 QY 61 TSINGVLWTVYHGAGTRTIAAPKGPVTQMTYTNVDKDLVGVQAPQGSRLTPTCTCGSSDLY 120
 DB 61 TSINGVLWTVYHGAGTRTIAAPKGPVTQMTYTNVDKDLVGVQAPQGSRLTPTCTCGSSDLY 120
 QY 121 LVTRHADVIPVRRGDSRGSLLSPRISYLVKSGSGGPLLCAGHAGVGFRAAVSTRGVAK 180
 DB 121 LVTRHADVIPVRRGDSRGSLLSPRISYLVKSGSGGPLLCAGHAGVGFRAAVSTRGVAK 180
 QY 181 AVDFIPVESLETTMRSP 197
 DB 181 AVDFIPVESLETTMRSP 197

RESULT 5
 AAB15221
 ID AAB15221 standard; protein: 197 AA.
 AC AAB15221;
 DT 19-DEC-2000 (first entry)
 DE Hepatitis C virus NS4A-NS3 fusion protease #3.
 DE
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
 KW liver failure; liver cancer; mutant; mutain.
 KW
 OS Hepatitis C virus.
 OS Synthetic.
 OS
 PN WO2000040707-A1.
 XX
 PD 13-JUL-2000.
 XX
 PF 06-JAN-2000; 2000WO-US00345.
 XX
 PR 08-JAN-1999; 99US-0115271.
 XX
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX
 PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
 XX WPI: 2000-465976/40.
 DR

DR WPI: 2000-465976/40.
 DR N-PSDB; AAA73330.
 XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 XX
 XX Claim 23; Fig 13; 66pp; English.
 PS
 CC The present sequence is a mutated version of a fusion protein created
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
 CC proteins are both essential for the replication of the virus, acting to
 CC cleave its replicative proteins from the polyprotein produced from the
 CC HCV genome. Inhibitors of the two proteins should be effective as
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A
 CC fusion proteins which can be used to identify inhibitors of this type, as
 CC well as enabling structural studies of the protease and
 CC protease:inhibitor complexes. This sequence contains the alpha-helix0-1
 CC variant.
 XX
 XX Sequence 197 AA;
 SQ
 Query Match 96.9%; Score 985; DB 21; Length 197;
 Best Local Similarity 97.0%; Pred. No. 3.2e-93;
 Matches 191; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 QY 1 MKKGSVVIVGRINLSGDTAYAQOTRGECCQETSGTGRDNKQVEGEVQIVSTATOTFLA 60
 DB 1 MKKGSVVIVGRINLSGDTAYAQOTRGECCQETSGTGRDNKQVEGEVQIVSTATOTFLA 60
 QY 61 TSINGVLTWTVYHGAGTRTIA SPKGPVTQMTYNDKDLVGMQAPQGSRSLLTPCTCGSSDLY 120
 DB 61 TCINGVLTWTVYHGAGTRTIA SPKGPVTQMTYNDKDLVGMQAPQGSRSLLTPCTCGSSDLY 120
 QY 121 LVTRHADVIPVRRRGRSGSLSPRISYLGSGGGPGLCPAGHAVGIFRAAVSTRGVAK 180
 DB 121 LVTRHADVIPVRRRGRSGSLSPRISYLGSGGGPGLCPAGHAVGIFRAAVSTRGVAK 180
 QY 181 AVDFIPVESLETTMRSP 197
 DB 181 AVDFIPVESLETTMRSP 197
 RESULT 6
 AAB15220
 ID AAB15220 standard; protein; 195 AA.
 XX
 AC AAB15220;
 DT 19-DEC-2000 (first entry)
 DE Hepatitis C virus NS4A-NS3 fusion protease #2.
 XX
 DE Hepatitis; NS3 protease; viral replication; chronic liver disease;
 KW liver failure; liver cancer; mutant; mutein.
 XX
 OS Hepatitis C virus.
 OS Synthetic.
 XX
 PN WO200040707-A1.
 XX
 PD 13-JUL-2000.
 XX
 PF 06-JAN-2000; 2000WO-US00345.
 XX
 PR 08-JAN-1999; 99US-0115271.
 XX
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX
 PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;

XX WPI: 2000-465976/40.
 DR N-PSDB; AAA73329.
 XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 XX
 XX Claim 23; Fig 12; 66pp; English.
 PS
 CC The present sequence is a mutated version of a fusion protein created
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
 CC proteins are both essential for the replication of the virus, acting to
 CC cleave its replicative proteins from the polyprotein produced from the
 CC HCV genome. Inhibitors of the two proteins should be effective as
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A
 CC fusion proteins which can be used to identify inhibitors of this type, as
 CC well as enabling structural studies of the protease and
 CC protease:inhibitor complexes. This sequence contains the alpha-helix0-1
 CC variant.
 XX
 XX Sequence 195 AA;
 SQ
 Query Match 93.5%; Score 951; DB 21; Length 195;
 Best Local Similarity 94.9%; Pred. No. 1e-89;
 Matches 187; Conservative 1; Mismatches 7; Indels 2; Gaps 1;
 QY 1 MKKGSVVIVGRINLSGDTAYAQOTRGECCQETSGTGRDNKQVEGEVQIVSTATOTFLA 60
 DB 1 MKKGSVVIVGRINLSGDTAYAQOTRGECCQETSGTGRDNKQVEGEVQIVSTATOTFLA 58
 QY 61 TSINGVLTWTVYHGAGTRTIA SPKGPVTQMTYNDKDLVGMQAPQGSRSLLTPCTCGSSDLY 120
 DB 59 TCINGVLTWTVYHGAGTRTIA SPKGPVTQMTYNDKDLVGMQAPQGSRSLLTPCTCGSSDLY 118
 QY 121 LVTRHADVIPVRRRGRSGSLSPRISYLGSGGGPGLCPAGHAVGIFRAAVSTRGVAK 180
 DB 119 LVTRHADVIPVRRRGRSGSLSPRISYLGSGGGPGLCPAGHAVGIFRAAVSTRGVAK 178
 QY 181 AVDFIPVESLETTMRSP 197
 DB 179 AVDFIPVESLETTMRSP 195
 RESULT 7
 AAB15226
 ID AAB15226 standard; protein; 197 AA.
 XX
 AC AAB15226;
 DT 19-DEC-2000 (first entry)
 DE Hepatitis C virus NS4A-NS3 fusion protease #8.
 XX
 DE Hepatitis; NS3 protease; viral replication; chronic liver disease;
 KW liver failure; liver cancer; mutant; mutein.
 XX
 OS Hepatitis C virus.
 OS Synthetic.
 XX
 PN WO200040707-A1.
 XX
 PD 13-JUL-2000.
 XX
 PF 06-JAN-2000; 2000WO-US00345.
 XX
 PR 08-JAN-1999; 99US-0115271.
 XX
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX

PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;

XX WPI: 2000-465976/40.

DR N-PSDB; AAA73335.

XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1

PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic

PT amino acid, useful for screening inhibitors that may treat hepatitis C

PT amino acid, useful for screening inhibitors that may treat hepatitis C

XX Example 5; Fig 18; 66pp; English.

PS The present sequence is a mutated version of a fusion protein created

XX using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These

CC proteins are both essential for the replication of the virus, acting to

CC cleave its replicative proteins from the polyprotein produced from the

CC HCV genome. Inhibitors of the two proteins should be effective as

CC antiviral treatments of HCV infection. This is useful as HCV can lead to

CC chronic liver disease such as cirrhosis, liver failure and liver cancer.

CC The present invention concerns a number of NS3 mutants and NS3-NS4A

CC fusion proteins which can be used to identify inhibitors of this type, as

CC well as enabling structural studies of the protease and

CC protease-inhibitor complexes. This sequence contains the alpha-helix0

CC wild-type sequence.

XX SQ Sequence 197 AA;

Query Match 93.0%; Score 946; DB 21; Length 197;

Best Local Similarity 94.4%; Pred. No. 3.3e-89;

Matches 186; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 MKKGSVVIVGRINLSGDTAYAAQOTRGEQCOETSGTRDKNOVEGEVQIVSTATQTFLA 60

DB 1 MKKGSVVIVGRINLSGDTAYAAQOTRGLGCIITSLTGRDKNOVEGEVQIVSTAQTFLA 60

QY 61 TSINGVLWTVYHAGTRTITASPKGPVTQMTYNDKDLVGMQAPQGSRLTPTCTCGSSDLY 120

DB 61 TCINGVCWTVYHAGTRTITASPKGPVTQMTYNDKDLVGMQAPQGSRLTPTCTCGSSDLY 120

QY 121 LVTRHADVIPVRRRGDSRGLSPRPISYLKSGSGPLLCPCPAGHAGVIFRAAVSTRGVAK 180

DB 121 LVTRHADVIPVRRRGDSRGLSPRPISYLKSGSGPLLCPCPAGHAGVIFRAAVSTRGVAK 180

QY 181 AVDFIPVESLETTMRSP 197

DB 181 AVDFIPVESLETTMRSP 197

RESULT 8

AA15212

ID AAB15212 standard; protein: 195 AA.

XX AC AAB15212;

XX AC AAB15212;

DT 19-DEC-2000 (first entry)

XX Hepatitis C virus NS4A-NS3 fusion protease #1.

DE Hepatitis C virus; single chain recombinant complex; linker;

XX Hepatitis; NS3 protease; viral replication; chronic liver disease;

KW liver failure; liver cancer.

XX Hepatitis C virus.

OS Synthetic.

XX WO2000040707-A1.

XX 13-JUL-2000.

XX 06-JAN-2000; 2000WO-US00345.

XX 08-JAN-1999; 99US-0115271.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

XX

PI

XX

DR

DR

XX

PT

PT

PT

XX

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

XX

SQ

Sequence 195 AA;

Query Match 89.7%; Score 912; DB 21; Length 195;

Best Local Similarity 92.4%; Pred. No. 1e-85;

Matches 182; Conservative 1; Mismatches 12; Indels 2; Gaps 1;

QY 1 MKKGSVVIVGRINLSGDTAYAAQOTRGEQCOETSGTRDKNOVEGEVQIVSTATQTFLA 60

DB 1 MKKGSVVIVGRINLSGDTAYAAQOTRGLGCIITSLTGRDKNOVEGEVQIVSTAQTFLA 58

QY 61 TSINGVLWTVYHAGTRTITASPKGPVTQMTYNDKDLVGMQAPQGSRLTPTCTCGSSDLY 120

DB 59 TCINGVCWTVYHAGTRTITASPKGPVTQMTYNDKDLVGMQAPQGSRLTPTCTCGSSDLY 118

QY 121 LVTRHADVIPVRRRGDSRGLSPRPISYLKSGSGPLLCPCPAGHAGVIFRAAVSTRGVAK 180

DB 119 LVTRHADVIPVRRRGDSRGLSPRPISYLKSGSGPLLCPCPAGHAGVIFRAAVSTRGVAK 178

QY 181 AVDFIPVESLETTMRSP 197

DB 179 AVDFIPVESLETTMRSP 195

RESULT 9

AA124943

ID AAY24943 standard; Protein: 665 AA.

XX AC AAY24943;

XX AC AAY24943;

DT 07-SEP-1999 (first entry)

XX HCV NS4A-NS3 complex SEQ ID NO:14.

DE HCV; hepatitis C virus; single chain recombinant complex; linker;

KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;

KW hydrophobic domain; covalent complex; detection; inhibitor.

XX Hepatitis C virus.

OS Synthetic.

XX WO928482-A2.

XX 10-JUN-1999.

XX 24-NOV-1998; 98WO-0524528.

XX 28-JUL-1998; 98US-0094331.

XX 28-NOV-1997; 97US-0067315.

XX PA (SCHE) SCHERING CORP.
 XX PI Malcolim BA, Taremi SS, Weber PC, Yao N;
 XX PI MPI; 1999-385385/32.
 XX DR New hepatitis C virus covalent complexes
 XX PT Claim 6; Page 90-92; 21lpp; English.
 XX PS The present invention describes a covalent hepatitis C virus (HCV)
 XX CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
 XX CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
 XX CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
 XX CC to the amino terminus of the HCV NS3 protease domain. The present
 XX CC sequence represents a specifically claimed example of the above
 XX CC complex. The covalent NS4A-NS3 complexes are useful for structural
 XX CC determination and determination of mode of binding of HCV inhibitors by
 XX CC NMR spectroscopy. They can also be used for detecting inhibitors of the
 XX CC protease activity, the helicase activity and the ATPase activity of NS3.
 XX CC The covalent NS4A-NS3 complexes are more soluble, stable and active than
 XX CC the non-covalent protease-peptide complexes previously available.
 XX SQ Sequence 665 AA;

Query Match 86.7%; Score 881.5; DB 20; Length 665;
 Best Local Similarity 85.2%; Pred. No. 8e-82;
 Matches 167; Conservative 16; Mismatches 10; Indels 3; Gaps 1;

QY 5 GSVVIVGRINLSGD---TAYAQOTRGEQCOETSGTRDKNOVEGEVQIVSTATQTFLAT 61
 DB 22 GSVVIVGRILSGSGSITAYSQOTRGLGCKKTSLTGRDKNOVEGEVQIVSTATQSFAT 81
 QY 62 SINGVLTMTVYHGAGTRTASPKGPVTOMYTNVDKDLVGMQAPGQSRSLPTCTCGSSDLYL 121
 DB 82 CVNGVCMTVYHGAGSKTLAGPKGPITOMYTNVDQDLVGMQAPPGARSLPTCTCGSSDLYL 141
 QY 122 VTRHADVIPVRRGDSRGSLSPRPISYLGSGGGPILCPAGHAVGIFRAAVSTRGVAKA 181
 DB 142 VTRHADVIPVRRGDSRGSLSPRPVSYLKGSGAGPILCPGSHAVGIFRAAVCTRGVAKA 201
 QY 182 VDFIPVESLETTMRSP 197
 DB 202 VDFVPVESMETTMRSP 217

RESULT 10
 AAY24947
 ID AAY24947 standard; Protein; 665 AA.
 XX AC AAY24947;
 XX DT 07-SEP-1999 (first entry)
 XX DE HCV NS4A-NS3 complex SEQ ID NO:18.
 XX KW HCV; hepatitis C virus; single chain recombinant complex; linker;
 XX KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
 XX KW hydrophobic domain; covalent complex; detection; inhibitor.
 XX OS Hepatitis C virus.
 XX OS Synthetic.
 XX PN WO9928482-A2.
 XX PD 10-JUN-1999.
 XX PF 24-NOV-1998; 98WO-US24528.
 XX PR 28-JUL-1998; 98US-0094331.
 XX PR 28-NOV-1997; 97US-0067315.

PA (SCHE) SCHERING CORP.
 XX PI Malcolim BA, Taremi SS, Weber PC, Yao N;
 XX PI MPI; 1999-385385/32.
 XX DR New hepatitis C virus covalent complexes
 XX PT Claim 6; Page 100-102; 21lpp; English.
 XX PS The present invention describes a covalent hepatitis C virus (HCV)
 XX CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
 XX CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
 XX CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
 XX CC to the amino terminus of the HCV NS3 protease domain. The present
 XX CC sequence represents a specifically claimed example of the above
 XX CC complex. The covalent NS4A-NS3 complexes are useful for structural
 XX CC determination and determination of mode of binding of HCV inhibitors by
 XX CC NMR spectroscopy. They can also be used for detecting inhibitors of the
 XX CC protease activity, the helicase activity and the ATPase activity of NS3.
 XX CC The covalent NS4A-NS3 complexes are more soluble, stable and active than
 XX CC the non-covalent protease-peptide complexes previously available.
 XX SQ Sequence 665 AA;

Query Match 86.4%; Score 878.5; DB 20; Length 665;
 Best Local Similarity 84.7%; Pred. No. 1.6e-81;
 Matches 166; Conservative 17; Mismatches 10; Indels 3; Gaps 1;

QY 5 GSVVIVGRINLSGD---TAYAQOTRGEQCOETSGTRDKNOVEGEVQIVSTATQTFLAT 61
 DB 22 GSVVIVGRILSGSGSITAYSQOTRGLGCKKTSLTGRDKNOVEGEVQIVSTATQSFAT 81
 QY 62 SINGVLTMTVYHGAGTRTASPKGPVTOMYTNVDKDLVGMQAPGQSRSLPTCTCGSSDLYL 121
 DB 82 CVNGVCMTVYHGAGSKTLAGPKGPITOMYTNVDQDLVGMQAPPGARSLPTCTCGSSDLYL 141
 QY 122 VTRHADVIPVRRGDSRGSLSPRPISYLGSGGGPILCPAGHAVGIFRAAVSTRGVAKA 181
 DB 142 VTRHADVIPVRRGDSRGSLSPRPVSYLKGSGAGPILCPGSHAVGIFRAAVCTRGVAKA 201
 QY 182 VDFIPVESLETTMRSP 197
 DB 202 VDFVPVESMETTMRSP 217

RESULT 11
 AAY24941
 ID AAY24941 standard; Protein; 665 AA.
 XX AC AAY24941;
 XX DT 07-SEP-1999 (first entry)
 XX DE HCV NS4A-NS3 complex SEQ ID NO:12.
 XX KW HCV; hepatitis C virus; single chain recombinant complex; linker;
 XX KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
 XX KW hydrophobic domain; covalent complex; detection; inhibitor.
 XX OS Hepatitis C virus.
 XX OS Synthetic.
 XX PN WO9928482-A2.
 XX PD 10-JUN-1999.
 XX PF 24-NOV-1998; 98WO-US24528.
 XX PR 28-JUL-1998; 98US-0094331.
 XX PR 28-NOV-1997; 97US-0067315.

(SCHE) SCHERING CORP.

XX
PI MalcolM BA, Taremi SS, Weber PC, Yao N;
XX WPI: 1999-385385/32.

DR
XX New hepatitis C virus covalent complexes

XX
XX Claim 6; Page 85-87; 21pp; English.

XX The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence represents a specifically claimed example of the above
CC complex. The covalent NS4A-NS3 complexes are useful for structural
CC determination and determination of mode of binding of HCV inhibitors by
CC NMR spectroscopy. They can also be used for detecting inhibitors of the
CC protease activity, the helicase activity and the ATPase activity of NS3.
CC The covalent NS4A-NS3 complexes are more soluble, stable and active than
CC the non-covalent protease-peptide complexes previously available.

XX
SQ Sequence 665 AA;

Query Match 86.3%; Score 877.5; DB 20; Length 665;
Best Local Similarity 85.2%; Pred. No. 2.1e-81;
Matches 167; Conservative 15; Mismatches 11; Indels 3; Gaps 1;

QY 5 GSVVIVGRINLSD---TAYAQOTRGECCQTSOTGRDKNQVEGEVQIVSTATOTFLAT 61
DB 22 GSVVIVGRILSGSITAYSQOTRGLGCKITSLTGRDKNQVEGEVQIVSTATQSFLAT 81
QY 62 SINGVLMTVYHGAGTRTIA SPKGPVTOMYTNVDKDLVGWQAPGQSRSLTPCTCGSSDLYL 121
DB 82 CVNGVCWTVYHGAGSKTLAGPKGPITOMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYL 141
QY 122 VTRHADVIPVRRGRDGRGSLSPRPISYLKSGSGGPLLCPCGHAVGIFRAAVSTRGVAKA 181
DB 142 VTRHADVIPVRRGRDGRGSLSPRPISYLKSGSGGPLLCPCGHAVGIFRAAVSTRGVAKA 201
QY 182 VDFIPVESLETTMRSP 197
DB 202 VDFVPVESMETTMRSP 217

RESULT 12

AAAY24942
ID AAY24942 standard; Protein; 665 AA.

XX
AC AAY24942;

XX
DT 07-SEP-1999 (first entry)

XX
DE HCV NS4A-NS3 complex SEQ ID NO:13.

XX HCV; hepatitis C virus; single chain recombinant complex; linker;
KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
KW hydrophobic domain; covalent complex; detection; inhibitor.

XX Hepatitis C virus.

OS Synthetic.

XX WO9928482-A2.

PN 10-JUN-1999.

PD 24-NOV-1998; 98WO-US24528.

PF 28-JUL-1998; 98US-0094331.

XX 28-NOV-1997; 97US-0067315.

PR (SCHE) SCHERING CORP.

XX
PA MalcolM BA, Taremi SS, Weber PC, Yao N;

XX

PI MalcolM BA, Taremi SS, Weber PC, Yao N;
XX WPI: 1999-385385/32.

DR
XX New hepatitis C virus covalent complexes

XX
XX Claim 6; Page 88-90; 21pp; English.

XX The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence represents a specifically claimed example of the above
CC complex. The covalent NS4A-NS3 complexes are useful for structural
CC determination and determination of mode of binding of HCV inhibitors by
CC NMR spectroscopy. They can also be used for detecting inhibitors of the
CC protease activity, the helicase activity and the ATPase activity of NS3.
CC The covalent NS4A-NS3 complexes are more soluble, stable and active than
CC the non-covalent protease-peptide complexes previously available.

XX
SQ Sequence 665 AA;

Query Match 86.3%; Score 877.5; DB 20; Length 665;
Best Local Similarity 85.2%; Pred. No. 2.1e-81;
Matches 167; Conservative 15; Mismatches 11; Indels 3; Gaps 1;

QY 5 GSVVIVGRINLSD---TAYAQOTRGECCQTSOTGRDKNQVEGEVQIVSTATOTFLAT 61
DB 22 GSVVIVGRILSGSITAYSQOTRGLGCKITSLTGRDKNQVEGEVQIVSTATQSFLAT 81
QY 62 SINGVLMTVYHGAGTRTIA SPKGPVTOMYTNVDKDLVGWQAPGQSRSLTPCTCGSSDLYL 121
DB 82 CVNGVCWTVYHGAGSKTLAGPKGPITOMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYL 141
QY 122 VTRHADVIPVRRGRDGRGSLSPRPISYLKSGSGGPLLCPCGHAVGIFRAAVSTRGVAKA 181
DB 142 VTRHADVIPVRRGRDGRGSLSPRPISYLKSGSGGPLLCPCGHAVGIFRAAVSTRGVAKA 201
QY 182 VDFIPVESLETTMRSP 197
DB 202 VDFVPVESMETTMRSP 217

RESULT 13

AAAY17880

ID AAY17880 standard; Protein; 216 AA.

XX
AC AAY17880;

XX
DT 07-SEP-1999 (first entry)

XX
DE HCV NS4A-NS3 complex SEQ ID NO:4.

XX HCV; hepatitis C virus; single chain recombinant complex; linker;
KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
KW hydrophobic domain; covalent complex; detection; inhibitor.

XX Hepatitis C virus.

OS Synthetic.

XX WO9928482-A2.

PN 10-JUN-1999.

PD 24-NOV-1998; 98WO-US24528.

PF 28-JUL-1998; 98US-0094331.

XX 28-NOV-1997; 97US-0067315.

PR (SCHE) SCHERING CORP.

XX
PA MalcolM BA, Taremi SS, Weber PC, Yao N;

XX

XX DR WPI: 1999-385385/32.
XX PT New hepatitis C virus covalent complexes
XX PS
XX Claim 6; Page 76-77; 21pp; English.
XX The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence represents a specifically claimed example of the above
CC complex. The covalent NS4A-NS3 complexes are useful for structural
CC determination and determination of mode of binding of HCV inhibitors by
CC NMR spectroscopy. They can also be used for detecting inhibitors of the
CC protease activity, the helicase activity and the ATPase activity of NS3.
CC The covalent NS4A-NS3 complexes are more soluble, stable and active than
CC the non-covalent protease-peptide complexes previously available.
XX SQ Sequence 216 AA;

Query Match 86.08; Score 874.5; DB 20; Length 216;
Best Local Similarity 85.18; Pred. No. 8.8e-82;
Matches 166; Conservative 16; Mismatches 10; Indels 3; Gaps 1;
QY 5 GSVVIVGRINLSGD---TAYAQOTRGEQCOETSGTGRDNQVEGEVQIVSTATQTFLAT 61
DB 22 GSVVIVGRILLSGSGSITAYSQOTRGLLCKKTSLTGRDNQVEGEVQIVSTATQSFAT 81
QY 62 SINGVLVTVYHGAGTRTIASPKGPVTOMYTNVDKLVGWQAPGSRSLTPTCTCGSSDLYL 121
DB 82 CVNGVCTVYHGAGSKTLAGPKGPITOMYTNVDQDLVGMQAPPGARSLTPTCTCGSSDLYL 141
QY 122 VTRHADVIPVRRRGDSRGSLLSPRPISYLKSGSGPLLCPCAGHAGVIFRAAVSTRGVAKA 181
DB 142 VTRHADVIPVRRRGDSRGSLLSPRPVSVYKSGAGPLLCPSGHAGVIFRAAVCTRGVAKA 201
QY 182 VDFIPVESLETTMRS 196
DB 202 VDFVPVESMETTMR 216

RESULT 14
RAY24945
ID AAY24945 standard; Protein: 665 AA.
XX AC AAY24945;
XX DT 07-SEP-1999 (first entry)
XX DE HCV NS4A-NS3 complex SEQ ID NO:16.

XX HCV; hepatitis C virus; single chain recombinant complex; linker;
KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
KW hydrophobic domain; covalent complex; detection; inhibitor.

OS Hepatitis C virus.
OS Synthetic.

XX WO9928482-A2.
XX PN 10-JUN-1999.
XX PD

XX PF 24-NOV-1998; 98WO-US24528.
XX XX
XX PR 28-JUL-1998; 98US-0094331.
XX PR 28-NOV-1997; 97US-0067315.

XX PA (SCHE) SCHERING CORP.

XX PI Malcolm BA, Taremi SS, Weber PC, Yao N;

XX WPI: 1999-385385/32.

DR WPI: 1999-385385/32.
XX PT New hepatitis C virus covalent complexes
XX PS
XX Claim 6; Page 95-97; 21pp; English.
XX The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence represents a specifically claimed example of the above
CC complex. The covalent NS4A-NS3 complexes are useful for structural
CC determination and determination of mode of binding of HCV inhibitors by
CC NMR spectroscopy. They can also be used for detecting inhibitors of the
CC protease activity, the helicase activity and the ATPase activity of NS3.
CC The covalent NS4A-NS3 complexes are more soluble, stable and active than
CC the non-covalent protease-peptide complexes previously available.
XX SQ Sequence 665 AA;

Query Match 86.08; Score 874.5; DB 20; Length 665;
Best Local Similarity 84.78; Pred. No. 4.2e-81;
Matches 166; Conservative 16; Mismatches 11; Indels 3; Gaps 1;
QY 5 GSVVIVGRINLSGD---TAYAQOTRGEQCOETSGTGRDNQVEGEVQIVSTATQTFLAT 61
DB 22 GSVVIVGRILLSGSGSITAYSQOTRGLLCKKTSLTGRDNQVEGEVQIVSTATQSFAT 81
QY 62 SINGVLVTVYHGAGTRTIASPKGPVTOMYTNVDKLVGWQAPGSRSLTPTCTCGSSDLYL 121
DB 82 CVNGVCTVYHGAGSKTLAGPKGPITOMYTNVDQDLVGMQAPPGARSLTPTCTCGSSDLYL 141
QY 122 VTRHADVIPVRRRGDSRGSLLSPRPISYLKSGSGPLLCPCAGHAGVIFRAAVSTRGVAKA 181
DB 142 VTRHADVIPVRRRGDSRGSLLSPRPVSVYKSGAGPLLCPSGHAGVIFRAAVCTRGVAKA 201
QY 182 VDFIPVESLETTMRS 197
DB 202 VDFVPVESMETTMR 217

RESULT 15
RAY24946
ID AAY24946 standard; Protein: 665 AA.
XX AC AAY24946;
XX DT 07-SEP-1999 (first entry)
XX DE HCV NS4A-NS3 complex SEQ ID NO:17.

XX HCV; hepatitis C virus; single chain recombinant complex; linker;
KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
KW hydrophobic domain; covalent complex; detection; inhibitor.

OS Hepatitis C virus.
OS Synthetic.

XX WO9928482-A2.
XX PN 10-JUN-1999.
XX PD

XX PF 24-NOV-1998; 98WO-US24528.
XX XX
XX PR 28-JUL-1998; 98US-0094331.
XX PR 28-NOV-1997; 97US-0067315.

XX PA (SCHE) SCHERING CORP.

XX PI Malcolm BA, Taremi SS, Weber PC, Yao N;

XX WPI: 1999-385385/32.

XX New hepatitis C virus covalent complexes
PT
XX
PS Claim 6; Page 97-99; 21lpp; English.
XX
CC The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence represents a specifically claimed example of the above
CC complex. The covalent NS4A-NS3 complexes are useful for structural
CC determination and determination of mode of binding of HCV inhibitors by
CC NMR spectroscopy. They can also be used for detecting inhibitors of the
CC protease activity, the helicase activity and the ATPase activity of NS3.
CC The covalent NS4A-NS3 complexes are more soluble, stable and active than
CC the non-covalent protease-peptide complexes previously available.
XX
SQ Sequence 665 AA:
Query Match 86.0%; Score 874.5; DB 20; Length 665;
Best Local Similarity 84.7%; Pred. No. 4.2e-81;
Matches 166; Conservative 16; Mismatches 11; Indels 3; Gaps 1;
QY 5 GSVVIVGRINLSGD---TAYAQOTRGEEGCOETISQTRGDKNQVEGEVQIVSTATQIFLAT 61
DB ||||| ||| |||:|||| || ||| ||||| |||||:|||||
22 GSVVIVGRILLSGSITAYSQOTRGLLGC1KTSLTGRDKNQVEGEVQIVSTATQSFAT 81
QY 52 SINGVLWTVYHGAGTRTIASPKGPVTOMYTNVDKDLVQWQAPQGSRLTPTCGSSDLYL 121
DB :||| ||||| |||||:|:| ||||| |||||:||||| ||||| |||||
82 CVNGVCWTVYHGAGSKTLAGPKGPITOMYTNVDQDLVQWQAPPGARSLTPTCGSSDLYL 141
QY 122 VTRHADVIPVRRRGRSGSLSPRPISYIKSGSGPLICPAGHAGVIFRAAVSTRGVAKA 181
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
142 VTRHADVIPVRRRGRSGSLSPRPVSYLKGAGGPLLCPSGHAGVIFRAAVCTRGVAKA 201
QY 182 VDFIPVESLETHRSP 197
DB |||:||||| |||||

Search completed: August 30, 2003, 19:12:24
Job time : 45.6227 secs

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	854.5	84.0	3011	1	GNWVC3	genome polyprotein
2	853.5	83.9	3011	1	S40770	genome polyprotein
3	848.5	83.4	3011	1	GNWVCH	genome polyprotein
4	837.5	82.4	3010	1	GNWVTW	genome polyprotein
5	827.5	81.4	3010	1	A45573	genome polyprotein
6	823.5	81.0	3010	1	GNWVCJ	genome polyprotein
7	823.5	81.0	3010	1	GNWVCJ	genome polyprotein
8	812.5	79.9	3010	1	SC630	genome polyprotein
9	743.5	73.1	3014	1	JS5620	genome polyprotein
10	675	66.4	3033	1	GNWVJ8	genome polyprotein
11	674	66.3	3033	1	JQ1303	genome polyprotein
12	249	24.5	3005	2	T08841	polyprotein - dour
13	243	23.9	2970	2	T08839	polyprotein - marm
14	92.5	9.1	590	2	B81104	nitrate/nitrite se
15	92.5	9.1	590	2	C81911	nitrate/nitrite se
16	85.5	8.4	209	2	H83144	probable aromatic
17	85.5	8.4	398	2	B71284	probable periplasm
18	83.5	8.2	1334	2	AB1775	hypothetical prote
19	83	8.2	452	2	I39383	angio-associated m
20	82.5	8.1	716	2	G83612	hypothetical prote
21	81	8.0	377	2	A75335	hypothetical prote
22	80	7.9	322	2	D87603	glycosyl transfera
23	80	7.9	477	2	E75392	hypothetical prote
24	80	7.9	915	2	F81196	transferrin-bindin
25	80	7.9	1615	2	JE0372	low density lipopr
26	79.5	7.8	514	2	AE2827	serine proteinase
27	79.5	7.8	2638	1	A42545	genome polyprotein
28	79	7.8	479	2	H70847	probable oxidoredu
29	79	7.8	3414	1	GNWVNE	genome polyprotein

```
Db 1005 RRGREILLGPADGMVSKGWRLLAPITAYAAQOTRGLGCIITSLTGRDKNOVEGEVQIVST 1064
QY 54 ATQTFLATSIINGVLTVYHGAGTRTIASPKGPVTOMYTNVDKDLVGVQAPQGSRLTPCT 113
Db 1065 AAOTFLATCINGVCWTVYHGAGTRTIASPKGPVIQMYTNVDQDLVGVWPAPOGSRSLTPCT 1124
QY 114 CGSSDLYLVTRHADVIPVRRGRDSRGLSPRPISYLVKSGSGGPLLCPAGHAGVIFRAAV 173
Db 1125 CGSSDLYLVTRHADVIPVRRGRDSRGLSPRPISYLVKSGSGGPLLCPAGHAGVIFRAAV 1184
QY 174 STRGVAKAVDFIPVESLETTMRSP 197
Db 1185 CTRGVAKAVDFIPVENLETTMRSP 1208

RESULT 2
S40770
genome polyprotein - hepatitis C virus
N:Contains: capsid protein C; envelope protein M; hepatitis virus (strain H)
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Date: 31-Dec-1992 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C:Accession: A36814; A41546
R:Inchausti, G.; Zebede, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
submitted to GenBank, July 1992
A:Description: Genomic structure of the human prototype strain H of hepatitis C virus
A:Reference number: A41546; MUID:92052256; PMID:1658800
A:Accession: A36814
A:Molecule type: genomic RNA
A:Residues: 1-3011 <INC>
A:Cross-references: GB:M67463; NID:g329737; PID:AAA45534.1; PID:g329738
R:Inchausti, G.; Zebede, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
A:Title: Genomic structure of the human prototype strain H of hepatitis C virus: comp
A:Reference number: A41546; MUID:92052256; PMID:1658800
A:Contents: annotation
A:Note: neither amino acid nor nucleotide sequence is given
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruct
F:116-191/Product: major envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis virus #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: DEXH motif
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240

Query Match 83.9%; Score 853.5; DB 1; Length 3011;
Best Local Similarity 82.8%; Pred. No. 3,4e-69;
Matches 169; Conservative 8; Mismatches 18; Indels 9; Gaps 1;
QY 3 KKGSVVIVGRIN-----LSGDTAYAAQOTRGEQCOET SQTGRDKNOVEGEVQIVST 53
Db 1005 RRGREILLGPADGMVSKGWRLLAPITAYAAQOTRGLGCIITSLTGRDKNOVEGEVQIVST 1064
QY 54 ATQTFLATSIINGVLTVYHGAGTRTIASPKGPVTOMYTNVDKDLVGVQAPQGSRLTPCT 113
Db 1065 AAOTFLATCINGVCWTVYHGAGTRTIASPKGPVIQMYTNVDQDLVGVWPAPOGSRSLTPCT 1124
QY 114 CGSSDLYLVTRHADVIPVRRGRDSRGLSPRPISYLVKSGSGGPLLCPAGHAGVIFRAAV 173
Db 1125 CGSSDLYLVTRHADVIPVRRGRDSRGLSPRPISYLVKSGSGGPLLCPAGHAGVIFRAAV 1184
QY 174 STRGVAKAVDFIPVESLETTMRSP 197
Db 1185 CTRGVAKAVDFIPVENLETTMRSP 1208

RESULT 3
GNMVTM
genome polyprotein - hepatitis C virus (strain H)
N:Contains: capsid protein C; envelope protein M; hepatitis virus (strain H)
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Date: 31-Dec-1992 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C:Accession: A36814; A41546
R:Inchausti, G.; Zebede, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
submitted to GenBank, July 1992
A:Description: Genomic structure of the human prototype strain H of hepatitis C virus
A:Reference number: A41546; MUID:92052256; PMID:1658800
A:Contents: annotation
A:Note: neither amino acid nor nucleotide sequence is given
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruct
F:116-191/Product: major envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis virus #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: DEXH motif
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240

Query Match 83.4%; Score 848.5; DB 1; Length 3011;
Best Local Similarity 81.9%; Pred. No. 9.7e-69;
Matches 167; Conservative 10; Mismatches 18; Indels 9; Gaps 1;
QY 3 KKGSVVIVGRIN-----LSGDTAYAAQOTRGEQCOET SQTGRDKNOVEGEVQIVST 53
Db 1005 RRGREILLGPADGMVSKGWRLLAPITAYAAQOTRGLGCIITSLTGRDKNOVEGEVQIVST 1064
QY 54 ATQTFLATSIINGVLTVYHGAGTRTIASPKGPVTOMYTNVDKDLVGVQAPQGSRLTPCT 113
Db 1065 AAOTFLATCINGVCWTVYHGAGTRTIASPKGPVIQMYTNVDQDLVGVWPAPOGSRSLTPCT 1124
QY 114 CGSSDLYLVTRHADVIPVRRGRDSRGLSPRPISYLVKSGSGGPLLCPAGHAGVIFRAAV 173
Db 1125 CGSSDLYLVTRHADVIPVRRGRDSRGLSPRPISYLVKSGSGGPLLCPAGHAGVIFRAAV 1184
QY 174 STRGVAKAVDFIPVESLETTMRSP 197
Db 1185 CTRGVAKAVDFIPVENLETTMRSP 1208

RESULT 4
GNMVTM
genome polyprotein - hepatitis C virus (strain Taiwan)
N:Contains: capsid protein C; envelope protein M; hepatitis virus (strain Taiwan)
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C:Accession: A40244
R:Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.
Virology 188, 102-113, 1992
```

```
Db 1005 RRGREILLGPADGMVSKGWRLLAPITAYAAQOTRGLGCIITSLTGRDKNOVEGEVQIVST 1064
QY 54 ATQTFLATSIINGVLTVYHGAGTRTIASPKGPVTOMYTNVDKDLVGVQAPQGSRLTPCT 113
Db 1065 AAOTFLATCINGVCWTVYHGAGTRTIASPKGPVIQMYTNVDQDLVGVWPAPOGSRSLTPCT 1124
QY 114 CGSSDLYLVTRHADVIPVRRGRDSRGLSPRPISYLVKSGSGGPLLCPAGHAGVIFRAAV 173
Db 1125 CGSSDLYLVTRHADVIPVRRGRDSRGLSPRPISYLVKSGSGGPLLCPAGHAGVIFRAAV 1184
QY 174 STRGVAKAVDFIPVESLETTMRSP 197
Db 1185 CTRGVAKAVDFIPVENLETTMRSP 1208

RESULT 2
S40770
genome polyprotein - hepatitis C virus
N:Contains: capsid protein C; envelope protein M; hepatitis virus (strain H)
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C:Accession: S40770; PC1285
R:Okamoto, H.
submitted to the EMBL Data Library, March 1992
A:Reference number: S40770
A:Accession: S40770
A:Molecule type: genomic RNA
A:Residues: 1-3011 <OKA>
A:Cross-references: EMBL:D10749; NID:g221586; PID:BA001582.1; PID:g221587
R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda,
Jpn. J. Exp. Med. 60, 167-177, 1990
A:Title: The 5'-terminal sequence of the hepatitis C virus genome.
A:Reference number: PC1284; MUID:91013116; PMID:2170712
A:Accession: PC1285
A:Molecule type: genomic RNA
A:Residues: 1-513 <OK>
A:Cross-references: GB:D00831; NID:g221511; PID:BA00705.1; PID:g221512
A:Experimental source: isolate HC-J1
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: major envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis virus #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 83.9%; Score 853.5; DB 1; Length 3011;
Best Local Similarity 82.8%; Pred. No. 3.4e-69;
Matches 169; Conservative 8; Mismatches 18; Indels 9; Gaps 1;
QY 3 KKGSVVIVGRIN-----LSGDTAYAAQOTRGEQCOET SQTGRDKNOVEGEVQIVST 53
Db 1005 RRGREILLGPADGMVSKGWRLLAPITAYAAQOTRGLGCIITSLTGRDKNOVEGEVQIVST 1064
QY 54 ATQTFLATSIINGVLTVYHGAGTRTIASPKGPVTOMYTNVDKDLVGVQAPQGSRLTPCT 113
Db 1065 AAOTFLATCINGVCWTVYHGAGTRTIASPKGPVIQMYTNVDQDLVGVWPAPOGSRSLTPCT 1124
QY 114 CGSSDLYLVTRHADVIPVRRGRDSRGLSPRPISYLVKSGSGGPLLCPAGHAGVIFRAAV 173
Db 1125 CGSSDLYLVTRHADVIPVRRGRDSRGLSPRPISYLVKSGSGGPLLCPAGHAGVIFRAAV 1184
QY 174 STRGVAKAVDFIPVESLETTMRSP 197
Db 1185 CTRGVAKAVDFIPVENLETTMRSP 1208
```


[illegible]

QY 174 STRGVAKAVDFIPVESLETTMRSP 197
 DB 1185 CTRGVAKAVDFIPVESMETTRSP 1208
 RESULT 7
 GNVVCJ
 genome polypeptide - hepatitis C virus (strain J)
 N:Contains: capsid protein C; envelope protein M; major envelope protein E; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 19-Jan-2001
 C:Accession: A39253; PS0086
 R:Kato, N.; Hijikata, M.; Otsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shimotoh
 Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990
 A:Title: Molecular cloning of the human hepatitis C virus genome from Japanese patients
 A:Reference number: A39253; MUID:91088550; PMID:2175903
 A:Accession: A39253
 A:Molecule type: genomic RNA
 A:Residues: 1-3010 <KAT>
 A:Cross-references: GB:D90208; MID:g221610; PIDN:BAAL4233.1; PID:g221611
 R:Kato, N.; Ohkoshi, S.; Shimotohno, K.
 Proc. Jpn. Acad. 65B, 219-223, 1989
 A:Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence vari
 A:Reference number: PS0086
 A:Accession: PS0086
 A:Molecule type: genomic RNA
 A:Residues: 2650-2707 <KAT>
 A:Experimental source: Japanese isolate
 C:Comment: The cleavage sites of this polypeptide have not been determined.
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serin
 F:116-191/Product: capsid protein C #status predicted <CPC>
 F:192-389/Product: major envelope protein E #status predicted <EPM>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>
 F:1230-1337/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B
 F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4a>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4b>
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,234,250,305,325,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2

A:Variety: isolate JKI
 C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 23-Mar-2001
 C:Accession: S18030; S33570; A48332; S18029
 R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
 submitted to the EMBL Data Library, September 1991
 A:Description: A whole genome of hepatitis C virus cDNA was isolated from a single pa
 A:Reference number: S18028
 A:Accession: S18030
 A:Molecule type: genomic RNA
 A:Residues: 1-3010 <HON>
 A:Cross-references: EMBL:X61596; MID:g59478; PIDN:CAA43793.1; PID:g59479
 A:Experimental source: isolate JKI from an individual
 R:Honda, M.; Kaneko, S.; Uonuma, M.; Kobayashi, K.; Murakami, S.
 Arch. Virol. 128, 163-169, 1993
 A:Title: Sequence analysis of putative structural regions of hepatitis C virus isolat
 A:Reference number: A48332; MUID:93119270; PMID:8380322
 A:Accession: S33570
 A:Molecule type: genomic RNA
 A:Residues: 1-547, 'T', 549-621, 'V', 623-624, 'S', 626-652, 'DL', 655-761, 'T', 763-782 <HON>
 A:Cross-references: EMBL:X61591
 A:Note: this sequence is inconsistent with the nucleotide translation
 A:Note: the authors translated the codon AGG for residue 43 as Pro, TGG for residue 3
 as Trp, and TTC for residue 771 as Ser
 A:Note: sequence extracted from NCBI backbone (NCBIN:121747, NCBI:121748)
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; se
 F:2-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <NS1>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>
 F:1230-1337/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B
 F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4a>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4b>
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,234,250,305,325,417,423,448,532,556,576,623,645/Binding site: carbohydrate
 Query Match 79.9%; Score 812.5; DB 1; Length 3010;
 Best Local Similarity 76.0%; Pred. No. 1.9e-65;
 Matches 155; Conservative 20; Mismatches 20; Indels 9; Gaps 1;
 QY 3 KGSVVIVGRIN-----LSGDTAYAQTRGEGCGQTSQTGRKNQVGEVQIVST 53
 DB 1005 RKREILLGPADSGEQWRLAPITAYSQQTRGLGCIITSLTGRKNQVGEVQIVST 1064
 QY 54 ATQTFLATISINGLVMTVYHGAGTRTIASPKGPVTOMTYNDKDLVGWQAPQSGRSRLTPCT 113
 DB 1065 ATQSLATCVNGVCVTVYHGAGSKTLAGPKGPINQMTNVDDDLVGWQAPSGAASLTPTCT 1124
 QY 114 CGSSDLYLTVTRHADVIPVRRGDSRGLSPRISYLKSGSGGPGLLCPAGHAGVIFRAAV 173
 DB 1125 YGSSDLYLTVTRHADVIPVRRGDSRGLSPRISYLKSGSGGPGLLCPAGHAGVIFRAAV 1184
 QY 174 STRGVAKAVDFIPVESLETTMRSP 197
 DB 1185 CTRGVAKAVDFIPVESMETTRSP 1208
 RESULT 9
 JC5620
 genome polypeptide - hepatitis C virus (isolate EUH1480)
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstru
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
 C:Accession: JC5620
 R:Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.
 Blochem. Biophys. Res. Commun. 236, 44-49, 1997
 A:Title: The complete coding sequence of hepatitis C virus genotype 5a, the predomina
 C:Reference number: JC5620; MUID:97366593; PMID:9223423

A:Accession: J05620
A:Molecule type: mRNA
A:Residues: 1-3014 <CHA>
A:Cross-references: GB:Y13184
A:Experimental source: genotype 5a, which predominates in South Africa
A:Note: the translation of the nucleotide sequence is not complete in this paper
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: major envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:384-406/Product: major envelope protein E #status predicted <MEE>
F:384-406/Region: hypervariable #status predicted
F:390-730/Product: nonstructural protein NS1 #status predicted <NS1>
F:731-1007/Product: nonstructural protein NS2 #status predicted <NS2>
F:1008-1616/Product: hepatitis C virus genome polyprotein
F:1231-1238/Region: nucleotide-binding motif A (P-loop)
F:1313-1318/Region: nucleotide-binding motif B
F:1317-1320/Region: DEXH motif
F:1617-1863/Product: nonstructural protein NS4a #status predicted <NS4a>
F:1864-2014/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>
F:2210-2249/Region: interferon sensitivity determining #status predicted

Query Match 73.1% Score 743.5; DB 1; Length 3014;
Best Local Similarity 68.6%; Pred. No. 3.8e-59;
Matches 140; Conservative 25; Mismatches 30; Indels 9; Gaps 1;

QY 3 KKGSVVIVGRIN-----LSGDTAYAQTRGEGCOETSGTGRDNQVEGVQIVST 53
DB 1006 RRGRIFGPADDIKTGWRLAPITAYAQTRGVGLVAIVLSLGRDNEAGEVQIFLT 1065
QY 54 ATQTFATSLVINGVLTVTHAGTRTASPKGPVTOMYTNVDKDLVGMWAPGSGSLTPTCT 113
DB 1066 ATQTFGLGICINGVMTTFHAGSKTLAGPKGPVQVOMYTNVDKDLVGMWAPGSGSLTCT 1125
QY 114 CGSSDLVLTVRHADYIPVRRGDSRGLSPRISYLAGSGGGLLCPAGHAGVGFRAAV 173
DB 1126 CGSADLVLTVRHADYIPARRGDTASLSRPSISYLAGSGGGLPIMCFSGHVGVRANV 1185
QY 174 STRGVAKAVDFIPVESLETTMRSP 197
DB 1186 CTRGVAKALEFVVENLETTMRSP 1209

RESULT 10
GNVJ38
genome polyprotein - hepatitis C virus (strain HC-J8)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C:Accession: A40250; PQ0397; PQ0559
R:Okamoto, H.; Kurai, K.; Okada, S.I.; Yamamoto, K.; Lizuka, H.; Tanaka, T.; Fukuda, S.; Virology 188, 331-341, 1992
A:Title: Full-length sequence of a hepatitis C virus genome having poor homology to rep
A:Reference number: A40250; MUID:92230232; PMID:1314459
A:Accession: A40250
A:Molecule type: genomic RNA
A:Residues: 1-3033 <OKA>
A:Cross-references: GB:D10988; GB:D01221; NID:g221608; PIDN:BAA01761.1; PID:g221609
R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L. J. Gen. Virol. 73, 1131-1141, 1992
A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e
A:Reference number: PQ0393; MUID:92268871; PMID:1316939
A:Accession: PQ0397
A:Molecule type: genomic RNA
A:Residues: 2678-2754 <CHA>
A:Cross-references: DBJ:D10134
A:Experimental source: isolate E-b12
R:Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Mori, S.; Hijikata, M.; Shimotohno Biochem. Biophys. Res. Commun. 181, 279-285, 1991
A:Title: Distribution of plural HCV types in Japan.
A:Reference number: PQ0554; MUID:92068204; PMID:1720309

A:Accession: PQ0559
A:Molecule type: mRNA
A:Residues: 2678-2729 <KAT>
A:Cross-references: GB:D10562; GB:D90518; NID:g221523; PIDN:BAA01418.1; PID:g221524
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruc
F:1-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: major envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-733/Product: nonstructural protein NS1 #status predicted <NS1>
F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>
F:1011-1619/Product: hepatitis C virus genome polyprotein
F:1234-1241/Region: nucleotide-binding motif A (P-loop)
F:1316-1321/Region: nucleotide-binding motif B
F:1320-1323/Region: DEXH motif
F:1620-1866/Product: nonstructural protein NS4a #status predicted <NS4a>
F:1867-2017/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,233,299,305,417,423,430,448,477,534,542,558,576,627,649,1091,1217,1259,203

Query Match 66.4% Score 675; DB 1; Length 3033;
Best Local Similarity 69.8%; Pred. No. 6.8e-53;
Matches 125; Conservative 24; Mismatches 30; Indels 0; Gaps 0;

QY 19 TAYAQTRGEGCOETSGTGRDNQVEGVQIVSTATQTFATSLVINGVLTVTHAGTRT 78
DB 1034 TAYTQTRGLGLVAIVLSLGRDNEAGQVQLSVSTQTFGLGTSISGLVLTVTHAGNKT 1093
QY 79 IASPKGPVTOMYTNVDKDLVGMWAPGSGSLTCTCGSSDLVLTVRHADYIPVRRGDSR 138
DB 1094 LAGPKGPVTOMYTSAGDGLVGMWAPGSGSLTCTCGAVDLVLTVRHADYIPVRRGDSR 1153
QY 139 GSLSRPSISYLAGSGGGLLCPAGHAGVGFRAAVSTRGVAKAVDFIPVESLETTMRSP 197
DB 1154 GALLSPRLSLTAGSGGGLVGMWAPGSGSLTCTCGAVDLVLTVRHADYIPVRRGDSR 1212

RESULT 11
JQ1303
genome polyprotein - hepatitis C virus (isolate HC-J6)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 17-Nov-2000
C:Accession: JQ1303
R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Kurai, K.; Iizuka, H.; Machida, A.; Miyakawa J. Gen. Virol. 72, 2697-2704, 1991
A:Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a
A:Reference number: JQ1303; MUID:92044440; PMID:1658196
A:Accession: JQ1303
A:Molecule type: genomic RNA
A:Residues: 1-3033 <OKA>
A:Cross-references: GB:D00944; NID:g221650; PIDN:BAA00792.1; PID:g221651
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; tr
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: major envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-733/Product: nonstructural protein NS1 #status predicted <NS1>
F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>
F:1011-1619/Product: hepatitis C virus genome polyprotein
F:1316-1321/Region: nucleotide-binding motif B
F:1320-1323/Region: DEXH motif
F:1620-1866/Product: nonstructural protein NS4a #status predicted <NS4a>
F:1867-2017/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,305,325,417,423,430,448,477,534,542,558,576,627,649,1091,1217,1259,203

Query Match 66.3% Score 674; DB 1; Length 3033;
Best Local Similarity 68.7%; Pred. No. 8.4e-53;
Matches 123; Conservative 27; Mismatches 29; Indels 0; Gaps 0;

QY 19 TAYAAQTRGEGGCGTQSGTGRKQNOVEGEQIVSTATQTFATSLNGVLVTVYHGAGTRT 78
 Db 1034 TAYAAQTRGGLCTIVSMTGRDTEQAGELOVLSTVTSQSLGTTISGLVTVYHGAGNKT 1093
 QY 79 IASPKGPVTOMYTNVDKDLVGMQAPQSGSLTPTCTCGSSDLVYLRHADVIPVRRRGDSR 138
 Db 1094 LAGSRGPVTOMYSSAEGDLVGMPSPPGTSLEPCTCGAVDLVYLRNADVIPARRRGDKR 1153
 QY 139 GSLSPRPISVLYKSGSGPLCPAGHAGVGFRAAVSTRGVAKAVDFIPVESLETTMRSP 197
 Db 1154 GALLSPRLSTLKGSSGPPVLCPRGHAGVGFRAAVSRGVAKSIDFIPVETLIDIVTRSP 1212
 RESULT 12
 T08841
 polypotein - douroucouli hepatitis GB virus A
 C:Species: douroucouli hepatitis GB virus A
 C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000
 C:Accession: T08841
 R:Erker, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.
 J. Gen. Virol. 79, 41-45, 1998
 A:Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.
 A:Reference number: 216486; MUID:98120818; PMID:9460920
 A:Accession: T08841
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-3005 <ERK>
 A:CROSS-references: EMBL:AF023425; NID:g2828599; PIDN:AAC40502.1; PID:g2828600
 C:Superfamily: hepatitis C virus genome polypotein
 C:Keywords: polypotein
 Query Match 24.5%; Score 249; DB 2; Length 3005;
 Best Local Similarity 33.5%; Pred. No. 5.4e-14;
 Matches 55; Conservative 30; Mismatches 69; Indels 10; Gaps 3;
 QY 33 ETSQTRDKNOVEGEQIVSTATQTFATSLNGVLVTVYHGAGTRTIASPKGPVTOMYTN 92
 Db 995 KISMGRDEREGESIVVLGTTSTRMCTCVNGVYITFHGNSNARTLAGVPVNCRWWS 1054
 QY 93 VDKLVGMQAPQSGSLTPTCTCGSSDLVYLRHADVIPVRRRGDSRGSLSPPISYLYK 152
 Db 1055 PSDVAVVYPLPSGASCLPCPKCGTQSVWCIRN--DGALCHGRLSKLVLDLPTAISDFRG 1112
 QY 153 SSGPGLPCPAGHAGVGFRAAVSTRGV-----AKAVDFIPVES 189
 Db 1113 SSGSPILCDGHHVGMH-VSVLHRGVKVTGVYRYKPKWETLPKDS 1155
 RESULT 13
 T08839
 polypotein - marmoset hepatitis GB virus A
 C:Species: marmoset hepatitis GB virus A
 C>Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 17-Nov-2000
 C:Accession: T08839
 R:Erker, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.
 J. Gen. Virol. 79, 41-45, 1998
 A:Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.
 A:Reference number: 216486; MUID:98120818; PMID:9460920
 A:Accession: T08839
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: genomic RNA
 A:Residues: 1-2970 <ERK>
 A:CROSS-references: EMBL:AF023424; NID:g2828597; PIDN:AAC40501.1; PID:g2828598
 C:Superfamily: hepatitis C virus genome polypotein
 C:Keywords: polypotein
 Query Match 23.9%; Score 243; DB 2; Length 2970;
 Best Local Similarity 27.4%; Pred. No. 1.9e-13;
 Matches 61; Conservative 40; Mismatches 80; Indels 42; Gaps 6;
 QY 3 KGSVVIVGRIN-----LSGDTAYAAQTRGEGCGTQSGTGRKQNOVEGEQIVS 52
 Db 946 RRGDEVILGVNLGWNELPPGFVPTAVPVVHHHGKGFVGKVTSMTGWDETEHVGNVVYL 1005

QY 53 TATOTFLATSLNGVLVTVYHGAGTRTIASPKGPVTOMYTNVDKDLVGMQAPQSGSLTPTC 112
 Db 1006 TSTRMCTCVNGVYITFHGNSNARTLAGVPVNCRWWSASDDVAVYPLPVGAKCLEPC 1065
 QY 113 TCGSSDLVYLRHADVIPVRRRGDSRGSLS-----PRPISYLYKSGSGPLCP 161
 Db 1066 KCQPOGVWVI-----RND--GALCHGTGRTVELDPAELCDFRSGSGSPILCD 1112
 QY 162 AGHAGVGFRAAVSTRG-----VAKAVDFIPVESLETTMRSP 197
 Db 1113 EGHAVGML-ISVLRHGSVTKIRYTKPWETLPREAITHTREAPP 1154
 RESULT 14
 B81104
 nitrate/nitrite sensor protein (EC 2.7.3.-) NMB1249 [similarity] - Neisseria meningitidis
 C:Species: Neisseria meningitidis
 C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
 C:Accession: B81104
 R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, R.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.; Qi, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.; Science 287, 1809-1815, 2000
 A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
 A:Reference number: A81000; MUID:20175755; PMID:10710307
 A:Accession: B81104
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-590 <TET>
 A:CROSS-references: GB:AE002473; GB:AE002098; NID:g7226488; PIDN:AAP41629.1; PID:g7722
 A:Experimental source: serogroup B, strain MC58
 C:Genetics:
 A:Gene: NMB1249
 C:Superfamily: nitrate/nitrite sensor protein narX
 C:Keywords: autophosphorylation; phosphohistidine; phosphoprotein; phosphotransferase
 F:95/Active site: His (phosphohistidine intermediate) #status predicted

Query Match 9.1%; Score 92.5; DB 2; Length 590;
 Best Local Similarity 21.3%; Pred. No. 1.5;
 Matches 46; Conservative 26; Mismatches 79; Indels 65; Gaps 6;
 QY 28 EBGQCTSGTGRDKNOVEGEQIVSTATQTFATSLNGVLVTVYHGAGTRTIASPKGPVT 87
 Db 213 EGGTPEFQVGRCFNMGGRKLKLYDDLEGQVAEQ-----TRSEKQONLNT 259
 QY 88 QMYTNVDKDLVGMQAPQ-----GSRSLTPTCTCGSSDLVYLRHAD----- 127
 Db 260 LLY-QTTRDLHQSIVTPOQAAEHFLNRLPAGVADSGRVCLDGGSDVTVSIHHDGCTAAS 318
 QY 128 -----VIPVRRGDSRGSLSPPISYLYKSGSGPLCPAGHAGVGFRAAVSTR---- 176
 Db 319 DLGKYHEEFPIEYQNETLGRLLSFPNGISLDEDRILLQTLGRLGVLGAKQEEK 378
 QY 177 -----GVAKAVDF--IPVESLET 192
 Db 379 RLAVLQERNLIAQGLHDSIAQALTFNLNQVQMLET 414

RESULT 15

C81911
 nitrate/nitrite sensor protein (EC 2.7.3.-) NMA1418 [similarity] - Neisseria meningitidis
 C:Species: Neisseria meningitidis
 C>Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
 C:Accession: C81911
 R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre Nature 404, 502-506, 2000
 A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491
 A:Reference number: A81775; MUID:20222556; PMID:10761919
 A:Accession: C81911
 A:Status: preliminary

A:Molecule type: DNA
A:Residues: 1-590 <PAR>
A:Cross-references: GB:AL162755; GB:AL157959; NID:g7379742; PIDN:CAB84658.1; PID:g738007
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: NMA1418
C:Superfamily: nitrate/nitrite sensor protein narX
C:Keywords: autophosphorylation; phosphohistidine; phosphoprotein; phosphotransferase; s
F:395/Active site: His (phosphohistidine intermediate) #status predicted

Query Match 9.18; Score 92.5; DB 2; Length 590;
Best local Similarity 21.3%, Pred. No. 1.5;
Matches 46; Conservative 26; Mismatches 79; Indels 65; Gaps 6;

QY 28 EGGQETQGRDNQVEGEVQIVSTATQTFLATSIINGVLMTVYHGAGTRTIASPKGPVT 87
DB 213 EGGTPEFKQVGRCFNQMGRLKILYDDLEGQVAEQ-----TRSLKQONLT 259
QY 88 QMYTNVDKDLVGMQAPQ-----GSRSLTPTCTGSSDLYLVTRHAD----- 127
DB 260 LLY-QTTRDLHQSYPQQAEEHFLNRLPAVGADSGRVCLDGGSDVYVSIHHADCGTAAS 318
QY 128 -----VIPVRRGDSRGSLLSPRPISYLYKSGSGPLLCPAGHAVGIFRAAVSTR--- 176
DB 319 DLGKYHEEIPLEYQNETPLGRLLLSFPNGISLDEDDRIQLTLGRQLGVSLAGAKQEEEX 378
QY 177 -----GVAKAVDF--IPVESLET 192
DB 379 RLLAVLQERNLIAQGLHDSIAQALTFLNLQVOMLET 414

Search completed: August 30, 2003, 19:20:30
Job time : 17.2134 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 18:01:52 ; Search time 9.75674 Seconds

(without alignments)
949.524 Million cell updates/sec

Title: US-09-965-594-18

Perfect score: 1017

Sequence: 1 MKKKGSVVIVGRINLSGDTA.....VAKAVDFIPVESLETTMRSP 197

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	854.5	84.0	3011	1	POLG_HCV1	P26664 h genome po
2	848.5	83.4	3011	1	POLG_HCVH	P27958 h genome po
3	837.5	82.4	3010	1	POLG_HCVTW	P29846 h genome po
4	827.5	81.4	3010	1	POLG_HCVTV	Q00269 h genome po
5	823.5	81.0	3010	1	POLG_HCVBK	P26663 h genome po
6	823.5	81.0	3010	1	POLG_HCVJA	P26662 h genome po
7	675	66.4	3033	1	POLG_HCVJ8	P26661 h genome po
8	674	66.3	3033	1	POLG_HCVJ6	P26660 h genome po
9	87	8.6	321	1	HROA_ARATH	Q98e17 arabidopsis
10	85.5	8.4	209	1	PAAD_PSEAE	Q9nx08 pseudomonas
11	83.5	8.2	437	1	DEGL_ARATH	Q22609 arabidopsis
12	83	8.2	452	1	AAMP_HUMAN	Q13685 homo sapien
13	79.5	7.8	3414	1	POLG_LANVT	P29837 l genome po
14	79	7.8	3414	1	POLG_TBVEW	P14336 t genome po
15	78.5	7.7	706	1	TRFE_HORSE	P27425 equus cabal
16	78.5	7.7	764	1	ICCR_DROME	Q08180 drosophila
17	78	7.7	911	1	TBIL_NEIMB	Q09056 neisseria m
18	78	7.7	3412	1	POLG_TBVEVS	P07720 t genome po
19	77.5	7.6	263	1	GRAK_MOUSE	Q35205 mus musculu
20	77	7.6	594	1	NIR_SPIOL	P05314 spinacia ol
21	77	7.6	1705	1	PPO_MOUSE	P70289 mus musculu
22	77	7.6	3414	1	POLG_TBVEH	Q01299 t genome po
23	76.5	7.5	323	1	VPKT_SMRVH	P21407 squirrel mo
24	76.5	7.5	333	1	MOSA_RHME	Q07607 rhizobium m
25	76.5	7.5	452	1	MLTD_ECOLI	P23931 escherichia
26	76	7.5	3411	1	POLG_YEFV1	P03314 y genome po
27	76	7.5	3411	1	POLG_YEFV2	P19901 y genome po
28	75.5	7.4	485	1	Y136_TREPA	O83172 treponema p
29	75.5	7.4	2269	1	WDR9_HUMAN	Q9ns16 homo sapien
30	75	7.4	467	1	NX1B_BOVIN	Q28142 bos taurus
31	75	7.4	973	1	VP18_HUMAN	Q9p253 homo sapien
32	75	7.4	1165	1	POL_GALV	P21414 gibbon ape
33	74.5	7.3	248	1	GRAD_MOUSE	P11033 mus musculu

RESULT 1

ID	POLG_HCV1	STANDARD;	PRT;	3011 AA.
AC	P26664;			
DT	01-AUG-1992 (Rel. 23, Created)			
DT	01-AUG-1992 (Rel. 23, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].			
DE	Hepatitis C virus (isolate 1) (HCV).			
OS	Hepatitis C virus positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.			
OC	Hepacivirus.			
OX	NCBI_TaxID=11104;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE-91172826; PubMed-1848704;			
RA	Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C., Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J., Bradley D.W., Kuo G., Houghton M.;			
RA	"Genetic organization and diversity of the hepatitis C virus.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).			
CC	-1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION. NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.			
CC	-1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.			
CC	-1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate + {RNA}(N).			
CC	-1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.			
CC	-1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL: M62321; AAA45676.1; -			
DR	PIR: A39166; GNWV3.			
DR	PDB: 1ALV; 16-FEB-99.			
DR	PDB: 1HEI; 25-NOV-98.			
DR	MEROPS; S29.001; -			
DR	InterPro; IPR001410; DEAD.			
DR	InterPro; IPR002522; HCV_capsid.			

P26596 lactococcus
P97608 rattus norv
P54748 rattus norv
P54856 ustilago ma
Q9eqs1 homo sapien
O88621 mus musculu
P08169 bos taurus
P28863 oryctolagus
Q28146 bos taurus
Q04538 t genome po
O67531 aquifex aeo
P34552 caenorhabdi

ALIGNMENTS

InterPro: IPR002521; HCV_core.
 InterPro: IPR002519; HCV_env.
 InterPro: IPR002531; HCV_NS1.
 InterPro: IPR002518; HCV_NS2.
 InterPro: IPR004109; HCV_NS3.
 InterPro: IPR000745; HCV_NS4.
 InterPro: IPR001490; HCV_NS4B.
 InterPro: IPR002868; HCV_NS5.
 InterPro: IPR002166; HCV_RdRp.
 InterPro: IPR001650; Helicase_C.
 InterPro: IPR007095; RNA_pol_DS_PS.
 InterPro: IPR007094; RNA_pol_PSVir.
 Pfam: PF01543; HCV_capsid; 1.
 Pfam: PF01542; HCV_core; 1.
 Pfam: PF01539; HCV_env; 1.
 Pfam: PF01560; HCV_NS1; 1.
 Pfam: PF01538; HCV_NS2; 1.
 Pfam: PF02907; HCV_NS3; 1.
 Pfam: PF01006; HCV_NS4a; 1.
 Pfam: PF01001; HCV_NS4b; 1.
 Pfam: PF01506; HCV_NS5a; 1.
 Pfam: PF00271; helicase_C; 1.
 Pfam: PF00998; Viral_RdRp; 1.
 Pfam: PD186062; HCV_NS1; 1.
 SMART: SM00487; DEXDc; 1.
 Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 3D-structure.
 INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 CELLULAR AMINOPEPTIDASE.
 CHAIN 1 115
 CHAIN 116 191
 CHAIN 192 383
 CHAIN 384 729
 CHAIN 730 1006
 CHAIN 1007 1615
 CHAIN 1616 1862
 CHAIN 1863 2013
 CHAIN 2014 3011
 CHAIN 3012 369
 TRANSMEM 347 369
 ACT_SITE 1083 1083
 ACT_SITE 1107 1107
 ACT_SITE 1165 1165
 NP_BIND 1230 1237
 SITE 1316 1319
 CARBOHYD 196 196
 CARBOHYD 209 209
 CARBOHYD 234 234
 CARBOHYD 305 305
 CARBOHYD 417 417
 CARBOHYD 423 423
 CARBOHYD 430 430
 CARBOHYD 448 448
 CARBOHYD 476 476
 CARBOHYD 532 532
 CARBOHYD 540 540
 CARBOHYD 556 556
 CARBOHYD 576 576
 CARBOHYD 623 623
 CARBOHYD 645 645
 CARBOHYD 2041 2041
 CARBOHYD 2077 2077
 CARBOHYD 2240 2240
 CARBOHYD 2364 2364
 CARBOHYD 2789 2789
 SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;
 Query Match 84.08; Score 854.5; DB 1; Length 3011;
 Best Local Similarity 82.88; Pred. No. 2.1e-71;
 Matches 169; Conservative 9; Mismatches 17; Indels 9; Gaps 1;

Db 1005 RRGREILLGADGMVSKGWRLLAPITAYAOQTGULLGCIITSLTGRDKNQVEGEIVST 1064
 Qy 54 ATQTFATSLNGVLVTVYHGAGTRTTIASPRGPTVQMTNTVDKDLVQWAPQGSRLTPTCT 113
 Db 1065 AAOTFLATCINGCVTVYHGAGTRTTIASPRGPTVQMTNTVDKDLVQWAPQGSRLTPTCT 1124
 Qy 114 CGSSDLYLVTRHADVIPIVRRGDSRGLSPRTISYLGSSGGPLLCPCGAGHAGVIFRAAV 173
 Db 1125 CGSSDLYLVTRHADVIPIVRRGDSRGLSPRTISYLGSSGGPLLCPCGAGHAGVIFRAAV 1184
 Qy 174 STRGVAKAVDFIPVESLETTMRSP 197
 Db 1185 CTRGVAKAVDFIPVENLETTMRSP 1208
 RESULT 2
 POLG_HCVH STANDARD; PRT; 3011 AA.
 AC P27958;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein p7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66); (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate H) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage: Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=11108;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92052256; PubMed=1658800;
 RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
 RA Prince A.M.;
 RT "Genomic structure of the human prototype strain H of hepatitis C
 RT virus: comparison with American and Japanese isolates.";
 RN Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
 RP [2]
 RX X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
 RA MEDLINE=97333322; PubMed=9187654;
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
 RT "Structure of the hepatitis C virus RNA helicase domain.";
 RN Nat. Struct. Biol. 4:463-467(1997).
 RP [3]
 RX X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
 RA MEDLINE=98154321; PubMed=9493270;
 RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
 RA Murcko M.A., Lin C., Caron P.R.;
 RT "Hepatitis C virus NS3 RNA helicase domain with a bound
 RT oligonucleotide: the crystal structure provides insights into the mode
 RT of unwinding.";
 RN Structure 6:89-100(1998).
 CC -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
 CC -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
 CC NS3-NS4A, NS4B-NS4B, NS4B-NS5A AND NS5A-NS5B.
 CC -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
 CC ACTIVATION OF NS3.
 CC -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
 CC -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the p6
 CC position, Cys or Thr in pi and Ser or Ala in pi',
 CC [RNA]{N}.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC [RNA]{N}.
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.

```

CC CC -1- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
CC CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
CC CC -1- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
CC CC -1- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL: M67463; AAA45334.1; .
DR DR PIR: A36814; GNMVCH.
DR DR PDB: 1HEI; 25-NOV-98.
DR DR PDB: 1AIV; 16-FEB-99.
DR DR PDB: 1A1R; 17-JUN-98.
DR DR MEROPS: S29.001; -.
DR DR MEROPS: U39.001; -.
DR DR TRANSFAC: T04155; -.
DR DR InterPro: IPR001410; DEAD.
DR DR InterPro: IPR002522; HCV capsid.
DR DR InterPro: IPR002521; HCV core.
DR DR InterPro: IPR002519; HCV env.
DR DR InterPro: IPR002531; HCV NS1.
DR DR InterPro: IPR002518; HCV NS2.
DR DR InterPro: IPR004109; HCV NS3.
DR DR InterPro: IPR000745; HCV NS4a.
DR DR InterPro: IPR001490; HCV NS4b.
DR DR InterPro: IPR002868; HCV NS5a.
DR DR InterPro: IPR002166; HCV RdRp.
DR DR InterPro: IPR001650; Helicase_C.
DR DR InterPro: IPR007095; RNA_pol_DS_Ps.
DR DR InterPro: IPR007094; RNA_pol_PSVir.
DR DR Pfam: PF01543; HCV_capsid; 1.
DR DR Pfam: PF01542; HCV_core; 1.
DR DR Pfam: PF01539; HCV_env; 1.
DR DR Pfam: PF01560; HCV_NS1; 1.
DR DR Pfam: PF01538; HCV_NS2; 1.
DR DR Pfam: PF02907; HCV_NS3; 1.
DR DR Pfam: PF01006; HCV_NS4a; 1.
DR DR Pfam: PF01001; HCV_NS4b; 1.
DR DR Pfam: PF01506; HCV_NS5a; 1.
DR DR Pfam: PF00271; helicase_C; 1.
DR DR Pfam: PF00998; Viral_RdRp; 1.
DR DR ProDom: PD186062; HCV_NS1; 1.
DR DR SMART: SM00487; DEXDC; 1.
KW KW Polypeptide: Glycoprotein; Transferrase; RNA-directed RNA polymerase;
KW KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW KW 3D-structure.
KW KW INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT FT CHAIN 1 191 CELLULAR AMINOPEPTIDASE.
FT FT CHAIN 192 383 CAPSID PROTEIN C.
FT FT CHAIN 384 746 ENVELOPE GLYCOPROTEIN E1.
FT FT CHAIN 747 809 ENVELOPE GLYCOPROTEIN E2.
FT FT CHAIN 810 1026 PROTEIN P7.
FT FT CHAIN 1027 1657 NONSTRUCTURAL PROTEIN NS2.
FT FT CHAIN 1658 1711 PROTEASE/HELICASE NS3.
FT FT CHAIN 1712 1972 NONSTRUCTURAL PROTEIN NS4a.
FT FT CHAIN 1973 2420 NONSTRUCTURAL PROTEIN NS4b.
FT FT CHAIN 2421 3011 NONSTRUCTURAL PROTEIN NS5a.
FT FT CHAIN 3011 369 NONSTRUCTURAL PROTEIN NS5b.
FT FT TRANSMEM 347 369 POTENTIAL.
FT FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT FT NP_BIND 1230 1237 SITE (POTENTIAL).
FT FT SITE 1316 1319 DECH BOX.
FT FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1224 1226 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1232 1233 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1236 1238 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1239 1246 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1247 1248 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1251 1255 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX 1258 1271 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1272 1272 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1277 1280 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1281 1282 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1283 1285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1291 1295 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1296 1301 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1302 1303 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1312 1316 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1317 1319 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX 1323 1335 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1336 1340 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1343 1347 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1352 1353 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1360 1361 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1362 1366 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1368 1368 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1373 1375 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1376 1377 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1378 1380 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX 1382 1385 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1389 1393 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX 1397 1409 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1410 1411 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1414 1417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1419 1420 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1432 1436 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1438 1439 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1450 1453 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1456 1463 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1471 1478 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1480 1480 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX 1481 1488 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1489 1490 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1497 1501 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1507 1507 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1511 1511 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX 1514 1527 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1532 1544 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1550 1550 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX 1555 1564 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX 1570 1578 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1579 1580 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX 1584 1597 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1598 1598 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX 1606 1611 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1614 1618 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1622 1623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1627 1627 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1635 1636 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX 1640 1652 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;
Query Match 83.4%; Score 848.5; DB 1; Length 3011;
Best Local Similarity 81.9%; Pred, No. 7.5e-71;

```


Matches 167; Conservative 10; Mismatches 18; Indels 9; Gaps 1;
QY 3 KKGSVIVGRIN-----LSGDTAYAQOTRGEQCOETSGTRKKNQVEGEVQIVST 53
DB 1005 RRQGEILLGADGKMGVSKGNFLAPITAYAAQTGRLGCGIITSILTRGXKNQVEGEVQIVST 1064
QY 54 ATQTFATSLNSVLTWVYHGAGTRTTASPKGPVQTYTWNVDKLVGWAQPOGSRSLTPCT 113
DB 1065 ATQTFATSLNSVLTWVYHGAGTRTTASPKGPVQTYTWNVDKLVGWAQPOGSRSLTPCT 1124
QY 114 CGSSDLYLVTRHADVTPVRRGDSRGLSPRISYLVKSSGGLLCPAGHAGVIFRAAV 173
DB 1125 CGSSDLYLVTRHADVTPVRRGDSRGLSPRISYLVKSSGGLLCPAGHAGVIFRAAV 1184
QY 174 STRGVAKAVDFIPVESLETTMRSP 197
DB 1185 CTRGVAKAVDFIPVENLETTMRSP 1208
RESULT 3
POLG_HCVTM STANDARD; PRT: 3010 AA.
AC P29846;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate Taiwan) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31645;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=92230206; PubMed=1314449;
RA Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;
RA "The Taiwanese hepatitis C virus genome: sequence determination and
RA mapping the 5' termini of viral genomic and antigenomic RNA.";
RL Virology 188:102-113(1992).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
CC RNA(N).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: M84754; ; NOT_ANNOTATED_CDS.
DR PIR: A40244; GNMVTH
DR PDB: 1N64; 25-FEB-03.
DR PDB: 1NS3; 08-APR-98.
DR MEROPS: S29.001; -.
DR MEROPS: U39.001; -.
DR InterPro: IPR001410; DEAD.

DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV env.
DR InterPro: IPR002531; HCV NS1.
DR InterPro: IPR002518; HCV NS2.
DR InterPro: IPR004109; HCV NS3.
DR InterPro: IPR000745; HCV NS4a.
DR InterPro: IPR001490; HCV NS4b.
DR InterPro: IPR002868; HCV NS5a.
DR InterPro: IPR002166; HCV RdRP.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV env; 1.
DR Pfam: PF01560; HCV NS1; 1.
DR Pfam: PF01538; HCV NS2; 1.
DR Pfam: PF02907; HCV NS3; 1.
DR Pfam: PF01006; HCV NS4a; 1.
DR Pfam: PF01001; HCV NS4b; 1.
DR Pfam: PF01506; HCV NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RdRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART; SMO0487; DEXDC; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 729
FT CHAIN 730 1006
FT CHAIN 1007 1615
FT CHAIN 1616 1862
FT CHAIN 1863 2013
FT CHAIN 2014 3010
FT TRANSMEM 347 369
FT ACT_SITE 1083 1083
FT ACT_SITE 1107 1107
FT ACT_SITE 1165 1165
FT SITE 1230 1237
FT SITE 1316 1319
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 233 233
FT CARBOHYD 234 234
FT CARBOHYD 250 250
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 532 532
FT CARBOHYD 540 540
FT CARBOHYD 556 556
FT CARBOHYD 576 576
FT CARBOHYD 623 623
FT CARBOHYD 645 645
FT CARBOHYD 2041 2041
FT CARBOHYD 2077 2077
FT CARBOHYD 2240 2240
FT CARBOHYD 2529 2529
FT CARBOHYD 2788 2788
SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDFE215 CRC64;
Query Match 82.4%; Score 837.5; DB 1; Length 3010;
Best Local Similarity 78.4%; Pred. No. 8e-70;
Matches 160; Conservative 18; Mismatches 17; Indels 9; Gaps 1;

Db	1005	RRGREIILGSPADSTEGOGWLLAPTAYAAQQRGLLGCIVTSLTGRDNKNQVEGEVQVNST	1064
QY	54	ATQTFLATISNGVLWTVYHGAGTRTIA SPKGPVTQMTYNTNVDKLDVWGQAOPGSRSLTPCT	1133
Db	1065	ATQSFSLATCVNGVCWTFVHGAGSKTLAGPKGPITQMTYNTVDQDLVGHWAHPGASRLTPCT	1124
QY	114	CGSDLVLTTHADVIVRRRGDSRGLSPRPISVTLKSGSGPLLPAGHANGVIGFRAAV	173
Db	1125	CGSDLVLTTHADVIVRRRGDSRGLSPRPISVTLKSGSGPLLPAGHANGVIGFRAAV	1184
QY	174	STRGVAKAVDFIPVESLETTMRSP	197
Db	1185	CTRGVAKAVDFIPVESMETMRSP	1208

RESULT 5

POLG_HCVBK	STANDARD;	PRT; 3010 AA.
ID	P26663;	
DT	01-AUG-1992 (Rel. 23, Created)	
DT	01-AUG-1992 (Rel. 23, Last sequence update)	
DT	15-SEP-2003 (Rel. 42, Last annotation update)	
DE	Genome polyprotein [Contents: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP33); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].	
OS	Hepatitis C virus (isolate BK) (HCV).	
OC	Viruses; ssRNA positive-strand viruses, no DNA stage: Flaviviridae; Hepacivirus.	
OX	NCBI_TaxId:11105;	
CC	[1]	
RN	SEQUENCE FROM N.A.	
RP	MEDLINE=91140698; PubMed=1847440;	
RX	Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J., Onishi E., Andoh T., Yoshida I., Okayama H.;	
RA	"Structure and organization of the hepatitis C virus genome isolated from human carriers.";	
RT	J. Virol. 65:1105-1113(1991).	
RL	[2]	
RN	SEQUENCE OF 1487-1500.	
RP	MEDLINE=96235224; PubMed=8647104;	
RX	Borowski P., Heiland M., Oehlmann K., Becker B., Kornetky L.;	
RA	"Non-structural protein 3 of hepatitis C virus inhibits phosphorylation mediated by cAMP-dependent protein kinase.";	
RT	Eur. J. Biochem. 237:611-618(1996).	
RL	[3]	
RN	X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.	
RP	MEDLINE=97015088; PubMed=8861916;	
RX	Love R.A., Padge H.E., Wickersham J.A., Hostomsky Z., Habuka N., Moosmaw E.W., Adachi T., Hostomsky Z.;	
RA	"The crystal structure of hepatitis C virus NS3 proteinase reveals a trypsin-like fold and a structural zinc binding site.";	
RT	Cell 87:331-342(1996).	
RL	[4]	
RN	X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.	
RP	MEDLINE=98227846; PubMed=9568891;	
RX	Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M., Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;	
RA	"Complex of NS3 protease and NS4A peptide of BK strain hepatitis C virus: a 2.2-A resolution structure in a hexagonal crystal form.";	
RT	Protein Sci. 7:837-847(1998).	
RL		
CC	-I- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.	
CC	NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.	
CC	-I- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the p6 position, Cys or Thr in Pl and Ser or Ala in Pl'.	
CC	-I- [CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate + [RNA](N).]	

FT CARBOHYD 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1031 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT HELIX 1039 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1050 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1059 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1068 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT TURN 1075 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1077 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT HELIX 1082 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT TURN 1086 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1090 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT TURN 1093 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1095 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT TURN 1101 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1104 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1108 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1120 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1122 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1129 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT TURN 1135 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1139 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1149 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT HELIX 1158 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT TURN 1162 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT TURN 1165 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1168 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT TURN 1172 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1175 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT TURN 1187 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1189 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT HELIX 1198 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT TURN 1203 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1680 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 3010 AA; F8422D5ECCDFD9C CRC64;

Query Match 81.0%; Score 823 5; DB 1; Length 3010;
 Best Local Similarity 76.5%; Pred. No. 1.6e-68;
 Matches 156; Conservative 21; Mismatches 18; Indels 9; Gaps 1;
 QY 3 KKGSVIVGRIN-----LSGDTAYAQQTRGEGCQETSGTGRDKNQVEGEVQIVST 53
 DB 1005 RRGKILLGPADSLRGLRLAPITAYSQTRGLGCIITSLTGRDKNQVEGEVQIVST 1064
 QY 54 ATQTEFLATSLNGVLTVYHAGTRTITASPKGPTVMYTNVDKLVGNQAPGSGSLPPT 113
 DB 1065 ATQSEFLATCVNGCVTVYHAGSKTTLAAPKGPITOMYTNVDQDLVGNPKPGARSLPPT 1124
 QY 114 CGSSDLYLVTRHADYIPVRRGDSRGLSLSPRISYLGKSSGGLLCPAGHAGVIFRAAV 173
 DB 1125 CGSSDLYLVTRHADYIPVRRGDSRGLSLSPRISYLGKSSGGLLCPFGHAGVIFRAAV 1184
 QY 174 STRGAKAVDFIPVESLETTMRSP 197
 DB 1185 CTRGAKAVDFIPVESMETTMRSP 1208

RESULT 6

POLG_HCVJA STANDARD; PRT; 3010 AA.
 AC P26662;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Envelope glycoprotein (Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (HCV).
 OS Hepatitis C virus (isolate Japanese) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91088550; PubMed=2175903;
 RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
 RA Sugimura T., Shimotohno K.;
 RT "Molecular cloning of the human hepatitis C virus genome from
 RT Japanese patients with non-A, non-B hepatitis";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
 RN [2]
 RP DISCUSSION OF SEQUENCE.
 RX MEDLINE=91192160; PubMed=1849488;
 RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraio K.,
 RA Ohkoshi S., Shimotohno K.;
 RT "Molecular structure of the Japanese hepatitis C viral genome.";
 RL FEBS Lett. 280:325-328(1991).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC {RNA}(N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: D90208; BA14233.1;
 CC PIR: A39253; GNMVCJ.
 CC HSP: P26663; LJXP.
 CC MEROPS: S29.001;
 CC MEROPS: U39.001;
 CC InterPro: IPR001410; DEAD.
 CC InterPro: IPR002522; HCV_capsid.
 CC InterPro: IPR002521; HCV_core.
 CC InterPro: IPR002519; HCV_env.
 CC InterPro: IPR002531; HCV_NS1.
 CC InterPro: IPR002518; HCV_NS2.
 CC InterPro: IPR004109; HCV_NS3.
 CC InterPro: IPR000745; HCV_NS4a.
 CC InterPro: IPR001490; HCV_NS4b.
 CC InterPro: IPR002868; HCV_NS5a.
 CC InterPro: IPR002166; HCV_RdRP.

DR InterPro: IPR001650; Helicase-C.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_capsid; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; Helicase-C; 1.
 DR Pfam: PF00998; Viral_RdRp; 1.
 DR Pfam: PF0186062; HCV_NS3; 1.
 DR Pfam: PF0186062; HCV_NS3; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 DR Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 DR Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
 DR INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT CHAIN 3010 369
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 532 532
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2788 2788
 FT SEQUENCE 3010 AA; 327017 MW; AA993794F460B185 CRC64;
 Query Match 81.08; Score 823.5; DB 1; Length 3010;
 Best Local Similarity 75.58; Pred. No. 1.6e-68;
 Matches 154; Conservative 23; Mismatches 18; Indels 9; Gaps 1;
 QY 3 KKGSVVIVGRINLSGD-----TAYAQTRGEGCQTSOTGRKNQVEGVIVST 53
 DB 1005 RRGKEILLGPADSFGEQGWRLAPITAYSQQTGRLGCIITSLTGRKNQVDEGVQLST 1064
 QY 54 ATQFLATISGVLTWYHAGTGTIAPKGPVTQNTYNDKDLVQWAPQGSRLTPTCT 113
 DB 1065 ATQSFALCVNGVCTWYHAGSAGTLAGKGPITQNTYNDQDLVQWAPPGARSTPTCT 1124
 QY 114 CGSSDLYLVTRHADVPVRRGDSRGSLSPRTSYLKGSGGGLPCPAGHAGVIFRAAV 173
 DB 1125 CGSSDLYLVTRHADVPVRRGDSRGSLSPRTSYLKGSGGGLPCPAGHAGVIFRAAV 1184
 QY 174 STRGVAKAVDFIPVESLETTMRSP 197

Db 1185 CTGKAVAKAVDFIPVESLETTMRSP 1208
 RESULT 7
 ID POLG_HCVJ8 STANDARD; PRT: 3033 AA.
 AC P26661;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4 (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-J8) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=11115;
 RN SEQUENCE FROM N.A.
 RP MEDLINE=92230232; PubMed=1314459;
 RA Okamoto H., Kurai K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,
 RA Fukuda S., Tsuda S., Mishiro S.;
 RT *Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes.*;
 RL Virology 188:331-341(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 CC {RNA}(N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D10988; BAA01761.1; -;
 CC PIR: A40250; GNMVJ8.
 CC HSP: P27958; 1HEI.
 CC MEROPS: S29.001; -;
 CC MEROPS: U39.001; -;
 CC InterPro: IPR001410; DEAD.
 CC InterPro: IPR002522; HCV_capsid.
 CC InterPro: IPR002521; HCV_core.
 CC InterPro: IPR002519; HCV_env.
 CC InterPro: IPR002531; HCV_NS1.
 CC InterPro: IPR002518; HCV_NS2.
 CC InterPro: IPR004109; HCV_NS3.
 CC InterPro: IPR000745; HCV_NS4a.
 CC InterPro: IPR001490; HCV_NS4b.
 CC InterPro: IPR002868; HCV_RdRp.
 CC InterPro: IPR002166; HCV_RSRP.
 CC InterPro: IPR007095; RNA_pol_DS_PS.
 CC InterPro: IPR007094; RNA_pol_PSVir.
 CC Pfam: PF01543; HCV_capsid; 1.
 CC Pfam: PF01542; HCV_core; 1.

DR Pfam: PF01539; HCV env; 1.
 DR Pfam: PF01560; HCV NS1; 1.
 DR Pfam: PF01538; HCV NS2; 1.
 DR Pfam: PF02907; HCV NS3; 1.
 DR Pfam: PF01006; HCV NS4a; 1.
 DR Pfam: PF01001; HCV NS4b; 1.
 DR Pfam: PF01506; HCV NS5a; 1.
 DR Pfam: PF00998; Viral_RdRp; 1.
 DR ProDom: PD18062; HCV NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 KW Polyprotein; RNA-directed RNA polymerase;
 KW Core protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural
 FT INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 233 233
 FT CARBOHYD 299 299
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CARBOHYD 627 627
 FT CARBOHYD 649 649
 FT CARBOHYD 1091 1091
 FT CARBOHYD 2038 2038
 FT CARBOHYD 2359 2359
 FT CARBOHYD 2811 2811
 FT SEQUENCE 3033 AA; 330177 MW; 1A173E7E3381FD1A CRC64;
 Query Match 66.4%; Score 675; DB 1; Length 3033;
 Best Local Similarity 69.8%; Pred. No. 1.3e-54;
 Matches 125; Conservative 24; Mismatches 30; Indels 0; Gaps 0;
 QY 19 TAYAAQTGEECCQTSQTDGKKNQVEQVIVSTATQTFATLSINGVLTWYHGAGTRT 78
 DB 1034 TAYTOOTRGLLGAIVYSLTGDRKNEQAGQVQLSVTQTGLTSGISGLVLTWYHGAGNKT 1093
 QY 79 IASPGKPTQMYNTVDKDLGVQAPQGSRSITPTCGSSDLYLVTRADVLPVRRRDSR 138
 DB 1094 LAGPKGPTQMYTSAGDLGVGSPGPGTKSLDPTCGNDVLYLVTRADVLPVRRKDDR 1153
 QY 139 GSELLSPRTSYLKGSGGLPLCLCPAGHAVGIFRAAVSTHGVAKAVDFIPVESLETHMRSP 197
 DB 1154 GALLSPRLSTLKGSGGPGVLCRSRHAVGLFRAAVCAARGVAKSIDFIPVESLDVATRTP 1212
 RESULT 8
 POLG_HCVJ6
 ID POLG_HCVJ6 STANDARD; PRT: 3033 AA.
 AC P26660;

DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DE 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus [isolate HC-J76] (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11113;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92044440; PubMed=1658196;
 RA Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H.,
 RA Machida A., Miyakawa Y., Mayumi M.;
 RT "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
 RT from a human carrier: comparison with reported isolates for conserved
 RT and divergent regions."; J. gen. Virol. 72:2697-2704(1991).
 RL J. gen. Virol. 72:2697-2704(1991).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 CC [RNA](N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MENA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D00944; BAA00792.1; -.
 CC PIR; JQ1303; JQ1303.
 CC HSSP; P27958; LHET.
 CC MEROPS; S29.001; -.
 CC MEROPS; U39.001; -.
 CC InterPro: IPR001410; DEAD.
 CC InterPro: IPR002522; HCV_capsid.
 CC InterPro: IPR002521; HCV_core.
 CC InterPro: IPR002519; HCV_env.
 CC InterPro: IPR002531; HCV_NSI.
 CC InterPro: IPR002518; HCV_NS2.
 CC InterPro: IPR004109; HCV_NS3.
 CC InterPro: IPR000745; HCV_NS4a.
 CC InterPro: IPR001490; HCV_NS4b.
 CC InterPro: IPR002868; HCV_NS5a.
 CC InterPro: IPR002166; HCV_RdRp.
 CC InterPro: IPR001650; Helicase_C.
 CC InterPro: IPR007095; RNA_pol_DS_PS.
 CC InterPro: IPR007094; RNA_pol_PSVir.
 CC Pfam: PF01543; HCV_capsid; 1.
 CC Pfam: PF01542; HCV_core; 1.
 CC Pfam: PF01539; HCV_env; 1.
 CC Pfam: PF01560; HCV_NSI; 1.
 CC Pfam: PF01338; HCV_NS2; 1.
 CC Pfam: PF02907; HCV_NS3; 1.
 CC Pfam: PF01006; HCV_NS4a; 1.
 CC Pfam: PF01001; HCV_NS4b; 1.
 CC Pfam: PF01506; HCV_NS5a; 1.

pfam: PF00271; helicase_C; 1.
 pfam: PF00998; Viral_RdRP; 1.
 ProDom: PD186062; HCV_NSL; 1.
 SMART: SMO0487; DEXDC; 1.
 Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 protein; Hydrolase; Serine protease.
 Transmembrane; Nonstructural
 REMOVED FROM CAPSID PROTEIN C BY THE
 INIT_MET 1
 CELLULAR AMINOPEPTIDASE
 CHAIN 1 115
 CHAIN 116 191
 CHAIN 192 383
 CHAIN 384 733
 CHAIN 734 1010
 CHAIN 1011 1619
 CHAIN 1620 1866
 CHAIN 1867 2017
 CHAIN 2018 3033
 CHAIN 3034 369
 TRANSMEM 347 369
 ACT_SITE 1087 1087
 ACT_SITE 1111 1111
 ACT_SITE 1169 1169
 NP_BIND 1234 1241
 SITE 1320 1323
 CARBOHYD 196 196
 CARBOHYD 209 209
 CARBOHYD 234 234
 CARBOHYD 305 305
 CARBOHYD 417 417
 CARBOHYD 423 423
 CARBOHYD 430 430
 CARBOHYD 448 448
 CARBOHYD 477 477
 CARBOHYD 534 534
 CARBOHYD 542 542
 CARBOHYD 558 558
 CARBOHYD 578 578
 CARBOHYD 627 627
 CARBOHYD 649 649
 CARBOHYD 1091 1091
 CARBOHYD 2038 2038
 CARBOHYD 2811 2811
 SEQUENCE 3033 AA; 329165 MW; F957F5C1A273BE9 CRC64;
 Query Match 66.3%; Score 674; DB 1; Length 3033;
 Best Local Similarity 68.7%; Pred. No. 1.6e-54;
 Matches 123; Conservative 27; Mismatches 29; Indels 0; Gaps 0;
 QY 19 TAYAQOTRGECCQETSGTGRKQVGEVQIVSTATQTFATLSINGVLVTVYHAGTRT 78
 DB 1034 TAYAQOTRGLGTVSGTGRKQVGEVQIVSTATQTFATLSINGVLVTVYHAGNKT 1093
 QY 79 TASPGRPVYMTYNDVLDVQWAPQGSRLPTCTCGSSLDLVTRHADVIPVRRRGDSR 138
 DB 1094 LAGSRGPVYMTYSSAEGDLVQWSPPTKSLPTCCGAVDLVTRNADVIPARRGDKR 1153
 QY 139 GLLSPRISVLKSSGGPLCPAGHAGVIFRAAVSTRGVAKAVDFIPVESLETTMRSP 197
 DB 1154 GALLSPRLSLTKSGSGGVLCPGRHAGVFRFAAVCSRGVAKSIDFIPVETLIDVTRSP 1212
 RESULT 9
 ID HHOA_ARATH STANDARD: PRT; 321 AA.
 AC Q9SEL7; O49507;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Protease HhoA, chloroplast precursor (EC 3.4.21.-).
 GN HHOA OR AT4G18370 OR F28J12.30.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

eurosids II: Brassicales; Brassicaceae; Arabidopsis.
 NCBI_TaxID=3702;
 [1]
 RP SEQUENCE FROM N.A.
 RA Lensch M.B.A., Sokolenko A., Herrmann R.G.;
 RT *Identification and characterization of the chloroplast RhoA protease,
 RT a homolog to the bacterial periplasmic protease RhoA*;
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 RN
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=20083488; PubMed=10617198;
 RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
 RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
 RA Harris B., Ansoorge W., Brandt P., Grivell L., Rieger M.,
 RA Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
 RA Kreis M., Delsens M., Puigdomenech P., Watson M., Schmidtheini T.,
 RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
 RA Vos P., Hohenseil J., Zimmermann W., Wedler H., Ridley P.,
 RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
 RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
 RA Braeken M., Weljens J., Voet M., Bastiaens I., Aert R., Defoor E.,
 RA Weitzenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
 RA Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
 RA Mooljman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
 RA Berner S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
 RA De Keyser A., Buysshaert C., Gielen J., Villarroel R., De Clercq R.,
 RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
 RA Clark L., Doggett J., Hall S., Kay M., Leonard N., McIlroy K., Mayes R.,
 RA Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
 RA Borkova D., Blocker H., Scharfe M., Grimm M., Loehner T.-H.,
 RA Dose S., de Haan M., Maarse A., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,
 RA Neumann S., Argirou A., Vitale D., Liquori R., Piravandi E.,
 RA Massenot O., Quigley F., Clabaud G., Muendlein A., Felber R.,
 RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
 RA Cheffor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
 RA Gibbons T., Weber N., Vandenbol M., Barques M., Terol J., Torres A.,
 RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
 RA Heijnen L., Haase D., Scholler P., Heber S., France P., Bielke C.,
 RA Frishman D., Schwarz S., Schoeller K., Mewes H.-W., Stocker S.,
 RA Zaccari P., Devan M., Wilson R.K., de la Bastide M., Habermann K.,
 RA Parnell I., Bedia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
 RA Stokem M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
 RA Stokem J., Kalicki J., Graves T., Harmon G., Edwards J.,
 RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
 RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
 RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
 RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
 RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
 RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
 RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
 RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
 RA Chen E., Marra M., Martienssen R., McCombie W.R.;
 RT *Sequence and analysis of chromosome 4 of the plant Arabidopsis
 RT thaliana*;
 RL Nature 402:769-777(1999).
 RN [3]
 RP SEQUENCE OF 72-82; 96-110; 150-159; 178-211 AND 306-320.
 RA Schubert M., Peterson U., Funk C., Haas B., Schroeder W.P.,
 RA Kieselbach T.;
 RT *The chloroplast lumen from Arabidopsis thaliana*;
 RL submitted (JUL-2001) to the SWISS-PROT data bank.
 CC -!- SUBCELLULAR LOCATION: Chloroplast; within the thylakoid lumen.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY 5C.
 CC -!- CAUTION: Ref.2 sequences differ from that shown due to erroneous
 CC gene model prediction. AT4G18370 and AT4G18375 were originally
 CC fused into a single gene.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

Db 296 GSVDCQAKLVSAATTGKVVGVFRPETVASQPSLGRGESESNVSL 341

RESULT 13

POID POLG_LANVT STANDARD: PRT: 3414 AA.

AC P29837:

DT 01-APR-1993 (Rel. 25, Created)

DT 01-FEB-1994 (Rel. 28, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Genome polyprotein [contains: Capsid protein C (Core protein);

DE Envelope protein PM; Matrix protein (Envelope protein M); Major

DE Envelope protein E; Nonstructural protein NS1; Nonstructural protein

DE NS2A; Nonstructural protein NS2B; Helicase/protease (EC 3.4.21.98)

DE (NS3); Nonstructural protein NS4A; Nonstructural protein NS4B; RNA-

DE directed RNA polymerase (EC 2.7.7.48) (NS5)].

OS Langkat virus (strain TP21).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Flavivirus.

OC NCBI_TaxID=31638;

RN [1]

RP SEQUENCE OF 1-776 FROM N.A.

RX MEDLINE=92074260; PubMed=1720591;

RA Mandl C.W., Iacono-Connors L., Wallner G., Holzmann H., Kunz C.,

RA Heinz F.X.,

RT *Sequence of the genes encoding the structural proteins of the low-

RT virulence tick-borne flaviviruses Langat TP21 and Yelantsev.*;

RN Virology 185:891-895(1991).

RN [2]

RP SEQUENCE OF 777-3414 FROM N.A.

RX MEDLINE=92263794; PubMed=1316884;

RA Iacono-Connors L.C., Schmaljohn C.S.;

RT *Cloning and sequence analysis of the genes encoding the

RT nonstructural proteins of Langkat virus and comparative analysis with

RT other flaviviruses.*;

RN Virology 188:875-880(1992).

CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral

CC precursor polyprotein, commonly with Asp or Glu in the P6

CC position, Cys or Thr in P1 and Ser or Ala in P1'.

CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +

CC (RNA)(N).

CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

CC PROTEIN C AND RNA.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: M73835; AAA02740.1; ALT_TERM.

DR EMBL: S35365; AAB22165.1; .

DR PIR: A42545; A42545.

DR HSSP: P14336; LSVB.

DR MEROPS: S07.001; .

DR InterPro: IPR001410; DEAD.

DR InterPro: IPR001122; Flavi_capsidC.

DR InterPro: IPR000336; Flavi_glycoproteE.

DR InterPro: IPR001850; Flavi_helicase.

DR InterPro: IPR000069; Flavi_M.

DR InterPro: IPR001157; Flavi_NS1.

DR InterPro: IPR000752; Flavi_NS2A.

DR InterPro: IPR000487; Flavi_NS2B.

DR InterPro: IPR000404; Flavi_NS4A.

DR InterPro: IPR001528; Flavi_NS4B.

DR InterPro: IPR000208; Flavi_NS5.

DR InterPro: IPR002535; Flavi_propep.

DR InterPro: IPR002877; FtsJ.

DR InterPro: IPR001650; Helicase_C.

DR InterPro: IPR007095; RNA_pol_DS.PS.

DR InterPro: IPR007094; RNA_pol_PSVir.

DR Pfam: PF01003; Flavi_capsid; 1.

DR Pfam: PF02832; Flavi_glycop_C; 1.

DR Pfam: PF00869; Flavi_glycoprot; 1.

DR Pfam: PF00949; Flavi_helicase; 1.

DR Pfam: PF01004; Flavi_M; 1.

DR Pfam: PF00948; Flavi_NS1; 1.

DR Pfam: PF01005; Flavi_NS2A; 1.

DR Pfam: PF01002; Flavi_NS2B; 1.

DR Pfam: PF01350; Flavi_NS4A; 1.

DR Pfam: PF01349; Flavi_NS4B; 1.

DR Pfam: PF00972; Flavi_NS5; 1.

DR Pfam: PF01570; Flavi_propep; 1.

DR Pfam: PF01728; FtsJ; 1.

DR Pfam: PF00271; helicase_C; 1.

DR ProDom: PD001556; Flavi_glycoproteE; 1.

DR ProDom: PD001496; Flavi_NS1; 1.

DR SMART: SM00487; DEXDC; 1.

DR SMART: SM00490; HELICC; 1.

KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;

KW Core protein; Coat protein; Envelope protein; Hydrolase; Helicase;

KW ATP-binding; Transmembrane; Nonstructural protein.

FT INIT_MET 1

CHAIN	1	112
FT CHAIN	113	205
FT CHAIN	206	280
FT CHAIN	281	776
FT CHAIN	777	1128
FT CHAIN	1129	1358
FT CHAIN	1359	1489
FT CHAIN	1490	2110
FT CHAIN	2111	2259
FT CHAIN	2260	2511
FT CHAIN	2512	3414

FT NP_BIND 1688 1695

FT SITE 1779 1782

FT TRANSMEM 103 119

FT TRANSMEM 262 278

FT TRANSMEM 728 744

FT TRANSMEM 758 774

FT DISULFID 283 310

FT DISULFID 340 396

FT DISULFID 354 385

FT DISULFID 372 401

FT DISULFID 466 570

FT DISULFID 587 618

FT CARBOHYD 144 144

FT CARBOHYD 434 434

SO SEQUENCE 3414 AA; 378017 MW; 59CB7E95DD70D82E CRC64;

Query Match 7.8%; Score 79.5; DB 1; Length 3414;

Best Local Similarity 22.7%; Pred. No. 69;

Matches 39; Conservative 21; Mismatches 67; Indels 45; Gaps 8;

QY 30 GCOETSGTGRDNQVEGEVQIVSTA-----TQTFLATSLNGVLWTVYH---GAGTRRTIAS 81

Db 1496 GCSEGRSDSRPLDVKNVGRVRYIYTPGLLMGQRIQVGYGAKGVLTMMHVTGAAALLVDGV 1555

QY 82 PKGPVTQMTYNDKDLV-----GMOA-----PGGSRSLTPTCTGSSDLVLT 123

Db 1556 AVGP---YWADVREDVVCYGGANSLERWRGRTVQVHAFPPG-RAHETHQCPGELLIL-- 1609

QY 124 RHADVIFVRRRGDSRGLSPRPISYLGKSGGGLCPAGHAGVIFRAAVST 175

Db 1610 -----ENGKMKAI----PIDLAKTSGSPIMNSGVEVGYGLNGLKT 1648

RESULT 14

POLG_TBEVW

ID POLG_TBEVW STANDARD; PRT: 3414 AA.


```
T DISULFID 498 692 BY SIMILARITY.
T DISULFID 508 522 BY SIMILARITY.
T DISULFID 519 533 BY SIMILARITY.
T DISULFID 590 604 BY SIMILARITY.
T DISULFID 642 647 BY SIMILARITY.
T METAL 79 79 IRON 1 (BY SIMILARITY).
T METAL 111 111 IRON 1 (BY SIMILARITY).
T METAL 209 209 IRON 1 (BY SIMILARITY).
T METAL 270 270 IRON 1 (BY SIMILARITY).
T METAL 413 413 IRON 2 (BY SIMILARITY).
T METAL 449 449 IRON 2 (BY SIMILARITY).
T METAL 544 544 IRON 2 (BY SIMILARITY).
T METAL 612 612 IRON 2 (BY SIMILARITY).
T BINDING 140 140 ANION (POTENTIAL).
T BINDING 480 480 ANION (POTENTIAL).
T CARBOHYD 515 515 N-LINKED (GLCNAC... ) (POTENTIAL).
T Q SEQUENCE 706 AA; 78094 MW; 1A0FA566C0409D8A CRC64;

Query Match 7.7%; Score 78.5; DB 1; Length 706;
Best Local Similarity 23.8%; Pred. No. 13;
Matches 48; Conservative 22; Mismatches 81; Indels 51; Gaps 11;

y 5 GSVTVVGRIMLSGDTAYAAQTRGEGCGQETSQTGRKKNQVEGEVQIVSTATQTFIATSIN 64
b 415 GFYIAGKCGLPVLAENYETRSACVDTPEGYH-----AVAVKSSSDPDLT---- 464

y 65 GVLWTVYHGAGTRTIAAPKGPVTQMTNVDKDLVGMQAPGSGRSRLTPCTCGSSDLYLVTR 124
b 465 ---WN-----SLKCK-KSCHTGVDR-TAGWNIPMGL-----LYSEIK 496

y 125 HADVIPRRRGDSRGLLSRPPISYLKGSGGP-LIC-PAGHA-----VGIPRAAVSTRG 177
b 497 HCEFDKFFREGCAPGYRRNSTLCNLICGSASGFGRECEPNNHERRYGYTGAFRCCLVERGD 556

y 178 VAKAVDFIPVESLE--TTMRSP 197
b 557 VA----FVKHQTVEQNTDGRNP 574
```

Search completed: August 30, 2003, 19:13:48
Job time : 11.7567 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 19:00:22 ; Search time 37.5921 Seconds
(without alignments)
1352.314 Million cell updates/sec

Title: US-09-965-594-18

Perfect score: 1017

Sequence: 1 MKKGSVVIVGRINLSGDTA.....VAKAVDFIPVESLETMRSP 197

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23.*

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mhc.*

8: sp_organelle.*

9: sp_phase.*

10: sp_plant.*

11: sp_rodent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

15: sp_rvirus.*

16: sp_bacteriaph.*

17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	872.5	85.8	4040	12	Q91FH8	Q91fh8 mucosal dis
2	858.5	84.4	3011	12	O36579	O36579 hepatitis c
3	854.5	84.0	2436	12	O81756	O81756 hepatitis c
4	854.5	84.0	3011	12	Q91FES	Q91fes hepatitis c
5	854.5	84.0	3011	12	Q91ELS8	Q91els8 hepatitis c
6	853.5	83.9	3011	12	O36508	O36508 hepatitis c
7	851.5	83.7	3011	12	O9PWX5	O9pwx5 hepatitis c
8	851.5	83.7	3015	12	O9PWU9	O9pwu9 hepatitis c
9	851.5	83.7	3015	12	Q9PWU9	Q9pwu9 hepatitis c
10	849	83.5	181	12	Q91RR8	Q91rr8 hepatitis c
11	849	83.5	181	12	Q91RT5	Q91rt5 hepatitis c
12	847	83.3	181	12	Q91RR5	Q91rr5 hepatitis c
13	847	83.3	181	12	Q91RR2	Q91rr2 hepatitis c
14	847	83.3	181	12	Q91RT9	Q91rt9 hepatitis c
15	846	83.2	181	12	Q91RR3	Q91rr3 hepatitis c
16	846	83.2	181	12	Q91RR4	Q91rr4 hepatitis c

17	846	83.2	181	12	Q91RS1	Q91rs1 hepatitis c
18	846	83.2	181	12	Q91RQ8	Q91rq8 hepatitis c
19	846	83.2	181	12	Q91RT1	Q91rt1 hepatitis c
20	846	83.2	181	12	Q91RR0	Q91rr0 hepatitis c
21	845.5	83.1	3011	12	O36609	O36609 hepatitis c
22	844	83.0	181	12	Q91RR6	Q91rr6 hepatitis c
23	844	83.0	181	12	Q91RS9	Q91rs9 hepatitis c
24	843	82.9	181	12	Q91RS3	Q91rs3 hepatitis c
25	842.5	82.8	3011	12	Q9DIT6	Q9dit6 hepatitis c
26	842	82.8	181	12	Q91RT4	Q91rt4 hepatitis c
27	842	82.8	181	12	Q91RS8	Q91rs8 hepatitis c
28	842	82.8	181	12	Q91RT3	Q91rt3 hepatitis c
29	842	82.8	181	12	Q91RS5	Q91rs5 hepatitis c
30	842	82.8	181	12	Q91RS7	Q91rs7 hepatitis c
31	842	82.8	181	12	Q91RT0	Q91rt0 hepatitis c
32	842	82.8	181	12	Q91RS2	Q91rs2 hepatitis c
33	841	82.7	181	12	Q91RS6	Q91rs6 hepatitis c
34	840.5	82.6	3010	12	O9QP61	O9qp61 hepatitis c
35	840	82.6	181	12	Q91RS4	Q91rs4 hepatitis c
36	839.5	82.5	3010	12	O68533	O68533 hepatitis c
37	839	82.5	181	12	Q91RR7	Q91rr7 hepatitis c
38	839	82.5	181	12	Q91RT6	Q91rt6 hepatitis c
39	839	82.5	3011	12	O36610	O36610 hepatitis c
40	838.5	82.4	361	12	O70817	O70817 hepatitis c
41	838	82.4	181	12	Q91RT8	Q91rt8 hepatitis c
42	837.5	82.4	361	12	O70818	O70818 hepatitis c
43	837	82.3	181	12	Q91RR9	Q91rr9 hepatitis c
44	836.5	82.3	3010	12	O9DTE2	O9dte2 hepatitis c
45	836.5	82.3	3010	12	O99AU2	O99au2 hepatitis c

ALIGNMENTS

RESULT 1

Q91FH8	PRELIMINARY;	PRT: 4040 AA.
ID	Q91FH8	
AC	Q91FH8;	
DT	01-OCT-2000 (Tremblrel. 15, Created)	
DT	01-OCT-2000 (Tremblrel. 15, Last sequence update)	
DT	01-MAR-2003 (Tremblrel. 23, Last annotation update)	
DE	Genome polyprotein.	
OS	Mucosal disease virus.	
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;	
OC	Pestivirus.	
OX	NCBI_TaxID=11099;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=20323484; PubMed=10864644;	
RA	Lai V.C., Zhong W., Skelton A., Ingravallo P., Vassilev V.,	
RA	Donis R.O., Hong Z., Lau J.Y.;	
RT	*Generation and characterization of a hepatitis C virus NS3 protease-	
RT	dependent bovine viral diarrhea virus.;	
RL	J. Virol. 74:6339-6347(2000).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RL	Lai V.C.H., Hong Z.;	
RL	Submitted (MAY-2000) to the EMBL/GenBank/DBDJ databases.	
DR	EMBL; AF268278; AAF82566.1; .	
DR	HSSP; P26863; IJXP.	
DR	MEROPS; S31.001; .	
DR	InterPro; IPR000280; CDvir_endptaseP80.	
DR	InterPro; IPR001410; DEAD.	
DR	InterPro; IPR004109; HCV NS3.	
DR	InterPro; IPR002166; HCV_RDRP.	
DR	InterPro; IPR001650; Helicase_C.	
DR	InterPro; IPR001005; Myb_DNA_Binding.	
DR	InterPro; IPR001568; RNase_T2.	
DR	InterPro; IPR007095; RNA_pol_DS_PS.	
DR	InterPro; IPR007094; RNA_pol_PSVir.	
DR	Pfam; PF02907; HCV_NS3; 1.	
DR	Pfam; PF00271; helicase_C; 1.	
DR	Pfam; PF00998; Viral_RDRP; 1.	

```

DR PRINTS: PR00729; COVENOPTASE.
DR SMART: SM00487; DEXDC; 1.
DR SMART: SM00490; HELICE; 1.
DR PROSITE: PS00037; MYB_1; 1.
DR PROSITE: PS00507; RDRP_POSITIVE; 1.
DR PROSITE: PS00521; RDRP_VIRAL; 1.
DR PROSITE: PS00531; RNASE_T2_2; 1.
DR ATP-binding; Helicase; Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transference.
SQ SEQUENCE 4040 AA; 453073 MW; ADE87791D05589DC CRC64;

Query Match 85.8%; Score 872.5; DB 12; Length 4040;
Best Local Similarity 89.2%; Pred. No. 1e-75;
Matches 174; Conservative 5; Mismatches 13; Indels 3; Gaps 1;

QY 5 GSVVIVGRINLSDG---TAYAAQOTRGEEGCOETSGTRDKKNQVEGEVQIVSTATQTFLAT 61
DB 10 GSVVIVGRIVLSSGSSITACAQOTRGLLGCKITSLTGRDKKNQVEGEVQIVSTATQTFLAT 69
QY 62 SINGVLTVYHGAGTRTIAISPKGPVTOMYTNVDKDLVGWQAPOGSRLTPTCTCGSSDLVYL 121
DB 70 CINGVCTVYHGAGTRTIAISPKGPVTOMYTNVDQDLVGPAPOGSRLTPTCTCGSSDLVYL 129
QY 122 VTRHADVIPVRRGDSRGSLSPRISYLYKSGSGGPLLCPAGHAVGIFRAAVSTRGVAKA 181
DB 130 VTRHANYIPVRRGDSRGSLSPRISYLYKSGSGGPLLCPAGHAVGLFRAAVCTRGVAKA 189
QY 182 VDFIPVESLETTRS 196
DB 190 VDFIPVENLETTRS 204

RESULT 2
Q36579 ID O36579 PRELIMINARY; PRT; 3011 AA.
AC O36579;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_taxid-11103;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-H77;
RX MEDLINE-97373636; PubMed-9228008;
RA Kolykhalov A.A., Agapov E.V., Blight K.J., Mihalik K., Feinstone S.M.,
RA Rice C.M.;
RT *Transmission of hepatitis C by intrahepatic inoculation with
RT transcribed RNA.*;
RL Science 277:570-574(1997).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MNA (BY SIMILARITY).
DR EMBL; AF009606; AAB66324.1; -.
DR HSSP; P27958; 1HEI.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR000745; HCV_NS4b.
DR InterPro; IPR001490; HCV_NS5a.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RDRP.
DR InterPro; IPR002166; HCV_RDRP.
DR InterPro; IPR001650; RNA_pol_DS_PS.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR InterPro; IPR007094; RNA_pol_PSVir.

DR PF01543; HCV_capsid; 1.
DR PF01542; HCV_core; 1.
DR PF01539; HCV_env; 1.
DR PF01560; HCV_NS1; 1.
DR PF01538; HCV_NS2; 1.
DR PF02907; HCV_NS3; 1.
DR PF01006; HCV_NS4a; 1.
DR PF01001; HCV_NS4b; 1.
DR PF01506; HCV_NS5a; 1.
DR PF00271; helicase_C; 1.
DR PF00998; Viral_RDRP; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS05057; RDRP_POSITIVE; 1.
DR PROSITE; PS05052; RDRP_VIRAL; 1.
KW ATP-binding; coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transference; Transmembrane.
SQ SEQUENCE 3011 AA; 327182 MW; E2E0E809C63C1B9 CRC64;

Query Match 84.4%; Score 858.5; DB 12; Length 3011;
Best Local Similarity 82.8%; Pred. No. 1.6e-74;
Matches 169; Conservative 10; Mismatches 16; Indels 9; Gaps 1;

QY 3 KGSVVIVGRIN-----LSGDTAYAAQOTRGEEGCOETSGTRDKKNQVEGEVQIVST 53
DB 1005 RRGQILLGPADGVMVSKGWRLAPITAYAAQOTRGLLGCIITSLTGRDKKNQVEGEVQIVST 1064
QY 54 ATQTFLATSLNGVLTVYHGAGTRTIAISPKGPVTOMYTNVDKDLVGWQAPOGSRLTPTCT 113
DB 1065 ATQTFLATSLNGVCTVYHGAGTRTIAISPKGPVTOMYTNVDQDLVGPAPOGSRLTPTCT 1124
QY 114 CGSSDLVLTVRHADVIPVRRGDSRGSLSPRISYLYKSGSGGPLLCPAGHAVGIFRAAV 173
DB 1125 CGSSDLVLTVRHADVIPVRRGDSRGSLSPRISYLYKSGSGGPLLCPAGHAVGLFRAAV 1184
QY 174 STRGVAKAVDFIPVESLETTRS 197
DB 1185 CTRGVAKAVDFIPVENLETTRS 1208

RESULT 3
Q81756 ID Q81756 PRELIMINARY; PRT; 2436 AA.
AC Q81756;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_taxid-11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Choo Q.-L., Richman K., Han J.;
RT *The nucleotide sequence of the Hepatitis C viral genome.*;
RL Submitted (MAY-1990) to the EMBL/GenBank/DBJ databases.
DR EMBL; M32084; AAA45677.1; -.
DR HSSP; P27958; 1AIV.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RDRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR PF01560; HCV_NS1; 1.
DR PF01538; HCV_NS2; 1.

```

```

DR Pfam: PF02907: HCV_NS3; 1.
DR Pfam: PF01006: HCV_NS4a; 1.
DR Pfam: PF01001: HCV_NS4b; 1.
DR Pfam: PF01306: HCV_NS5a; 1.
DR Pfam: PF00271: helicase.C; 1.
DR Pfam: PF00998: HCV_RDRP; 1.
DR ProDom: PD186062: HCV_NSL; 1.
DR SMART: SM00487: DEXDC; 1.
DR PROSITE: PS0507: RDRP_POSITIVE; 1.
DR PROSITE: PS0521: RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolyase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
FT NON_TER 1
FT TER 2436
SQ SEQUENCE 2436 AA; 264734 MW; D7B9872900BE3125 CRC64;

Query Match 84.0%; Score 854.5; DB 12; Length 2436;
Best Local Similarity 82.8%; Pred. No. 2.9e-74;
Matches 169; Conservative 9; Mismatches 17; Indels 9; Gaps 1;

QY 3 KGSVWIVGRIN-----LSGDTAYAAQTRGEGCGQETSGTGRDKNQVEGEVQIVST 53
Db 555 RRGREILLGPADGWSKGRLLAPITAYAAQTRGLGCIITSLTGRDKNQVEGEVQIVST 614
QY 54 ATOTFLATISNGVLTVYHGAGTRTIASPKGPVTOMYTNVDKLVGWAQPGSRLTPTCT 113
Db 615 AAQTFLATCINGVCWTYHGAGTRTIASPKGPVQIOMYTNVDKLVGWAQPGSRLTPTCT 674
QY 114 CGSSDLYLVTRHADVLPVRRGDSRGLSPRPISYLYKSGSGGLPCPAGHAGVIFRAAV 173
Db 675 CGSSDLYLVTRHADVLPVRRGDSRGLSPRPISYLYKSGSGGLPCPAGHAGVIFRAAV 734
QY 174 STRGVAKAVDFIPVESLETHMRSP 197
Db 735 CTRGVAKAVDFIPVENLETHMRSP 758

RESULT 4
ID Q9IF55 PRELIMINARY; PRT; 3011 AA.
AC Q9IF55;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=21262212; PubMed=11369872;
RA Lanford R.E., Lee H., Chavez D., Guerra B., Brasky K.M.;
FT "Infectious cDNA clone of the hepatitis C virus genotype 1 prototype
sequence."
RL J. Gen. Virol. 82:1291-1297(2001).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AF271632; AAF81759.1; -.
DR HSSP; P27958; 1A1V.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_NSL.
DR InterPro; IPR002531; HCV_NSL1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.

InterPro: IPR002868; HCV_NS5a.
InterPro: IPR002166; HCV_RDRP.
InterPro: IPR001650; Helicase.C.
InterPro: IPR007095; RNA_pol_DS_PS.
InterPro: IPR007094; RNA_pol_PSVir.
Pfam: PF01543; HCV_capsid; 1.
Pfam: PF01542; HCV_core; 1.
Pfam: PF01539; HCV_env; 1.
Pfam: PF01560; HCV_NSL; 1.
Pfam: PF01538; HCV_NSL1; 1.
Pfam: PF02907; HCV_NS3; 1.
Pfam: PF01006; HCV_NS4a; 1.
Pfam: PF01001; HCV_NS4b; 1.
Pfam: PF01506; HCV_NS5a; 1.
Pfam: PF00271; helicase.C; 1.
Pfam: PF00998; Viral_RDRP; 1.
ProDom: PD186062; HCV_NSL; 1.
SMART: SM00487; DEXDC; 1.
PROSITE: PS00190; CYTOCHROME_C; 1.
PROSITE: PS0507; RDRP_POSITIVE; 1.
PROSITE: PS0521; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolyase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA; 327124 MW; 2489CE74AC864E58 CRC64;

Query Match 84.0%; Score 854.5; DB 12; Length 3011;
Best Local Similarity 82.8%; Pred. No. 3.9e-74;
Matches 169; Conservative 9; Mismatches 17; Indels 9; Gaps 1;

QY 3 KGSVWIVGRIN-----LSGDTAYAAQTRGEGCGQETSGTGRDKNQVEGEVQIVST 53
Db 1005 RRGREILLGPADGWSKGRLLAPITAYAAQTRGLGCIITSLTGRDKNQVEGEVQIVST 1064
QY 54 ATOTFLATISNGVLTVYHGAGTRTIASPKGPVTOMYTNVDKLVGWAQPGSRLTPTCT 113
Db 1065 AAQTFLATCINGVCWTYHGAGTRTIASPKGPVQIOMYTNVDKLVGWAQPGSRLTPTCT 1124
QY 114 CGSSDLYLVTRHADVLPVRRGDSRGLSPRPISYLYKSGSGGLPCPAGHAGVIFRAAV 173
Db 1125 CGSSDLYLVTRHADVLPVRRGDSRGLSPRPISYLYKSGSGGLPCPAGHAGVIFRAAV 1184
QY 174 STRGVAKAVDFIPVESLETHMRSP 197
Db 1185 CTRGVAKAVDFIPVENLETHMRSP 1208

RESULT 5
ID Q9ELS8 PRELIMINARY; PRT; 3011 AA.
AC Q9ELS8;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Desai S.M., Devare S., Yamaguchi J.;
RT "Hepatitis C Virus."
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AF290978; AAG02099.1; -.
DR HSSP; P27958; 1HEI.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR001410; DEAD.

```


DR InterPro: IPR002522; HCV capsid.
 DR InterPro: IPR002521; HCV core.
 DR InterPro: IPR002519; HCV env.
 DR InterPro: IPR002531; HCV NS1.
 DR InterPro: IPR002518; HCV NS2.
 DR InterPro: IPR004109; HCV NS3.
 DR InterPro: IPR000745; HCV NS4a.
 DR InterPro: IPR001490; HCV NS4b.
 DR InterPro: IPR002868; HCV NS5a.
 DR InterPro: IPR002166; HCV NS5a.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_Psvir.
 DR Pfam: PF01543; HCV core; 1.
 DR Pfam: PF01542; HCV core; 1.
 DR Pfam: PF01539; HCV env; 1.
 DR Pfam: PF01560; HCV NS1; 1.
 DR Pfam: PF01538; HCV NS2; 1.
 DR Pfam: PF02907; HCV NS3; 1.
 DR Pfam: PF01006; HCV NS4a; 1.
 DR Pfam: PF01001; HCV NS4b; 1.
 DR Pfam: PF01506; HCV NS5a; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00998; Viral_Rdrp; 1.
 DR ProDom: PD186062; HCV NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR PROSITE: PS00190; CYTOCHROME_C; 1.
 DR PROSITE: PS05057; RDRP_POSITIVE; 1.
 DR PROSITE: PS05021; RDRP_VIRAL; 1.
 KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
 KW Hydrolyase; Nonstructural protein; Polyprotein;
 KW RNA-directed RNA polymerase; transferase; transmembrane.
 SQ SEQUENCE 3011 AA; 327107 MW; A6BECF9A3BEE13F CRC64;

Query Local 84.0%; Score 854.5; DB 12; Length 3011;
 Best Local Similarity 82.4%; Pred. No. 3.9e-74;
 Matches 168; Conservative 11; Mismatches 16; Indels 9; Gaps 1;

QY 3 KGSVVIVGRIN-----LSGDTAYAQTRGEGCOETSGTRDKNQVEGEVQIVST 53
 DB 1005 RRGQELIPADGMVSKOMRLLAPITAYAQTRGLGCIITSLGRDKNQVEGEVQIVST 1064

QY 54 ATOTFLATISINGLVTVHAGTRTIASPKGPVTOMTVNDKLVGQAPGSGSLTPCT 113
 DB 1065 ATQTLATCINGCVTVHAGTRTIASPKGPVTOMTVNDKLVGQAPGSGSLTPCT 1124

QY 114 CGSSDLVLVTRHADVIPRRGDSRGLSLSPRISYLGSGGGLPCPAGHAVGIFRAAV 173
 DB 1125 CGSSDLVLVTRHADVIPRRGDSRGLSLSPRISYLGSGGGLPCPAGHAVGIFRAAV 1184

QY 174 STRGVAKAVDFIPVESLTTMRSP 197
 DB 1185 CTRGVAKAVDFIPVENLETHRSP 1208

RESULT 6
 Q03463 PRELIMINARY; PRT; 3011 AA.
 ID Q03463
 AC Q03463;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Genome polyprotein.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_Taxid=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-HC-J1;
 RX MEDLINE-91013116; PubMed-2170712;
 RA Okamoto H., Okada S., Sugiyama Y., Yotsumoto S., Tanaka T.,
 Yoshizawa H.;

RT The 5'-terminal sequence of the hepatitis C virus genome.*;
 RL Jpn. J. Exp. Med. 60:167-177(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-HC-J1;
 RX MEDLINE-9204440; PubMed-1658196;
 RA Okamoto H., Okada S., Sugiyama Y., Kurai K., Iizuka H., Machida A.,
 Miyakawa Y., Mayumi M.;
 RA *Nucleotide sequences of the genomic RNA of hepatitis C virus isolated
 RT from a human carrier: comparison with reported isolates for conserved
 RT and divergent regions.*;
 RL J. Gen. Virol. 72:2697-2704(1991).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-HC-J1;
 RX MEDLINE-93117120; PubMed-1335573;
 RA Okamoto H., Kanai N., Mishiro S.;
 RT *Full-length nucleotide sequence of a Japanese hepatitis C virus
 RT isolate (HC-J1) with high homology to USA isolates.*;
 RL Nucleic Acids Res. 20:6410-6410(1992).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-HC-J1;
 RA Okamoto H.;
 RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN-HC-J1;
 RX MEDLINE-94174722; PubMed-7510436;
 RA Mink M., Benichou S., Madaule P., Tiollais P., Prince A.,
 Inchausti G.;
 RA *Characterization and mapping of a B-cell immunogenic domain in
 RT hepatitis C virus E2 glycoprotein using a yeast peptide library.*;
 RL Virology 200:246-255(1994).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 DR EMBL: D10749; BAA01582.1; -;
 DR HSSP: P27958; 1HEI.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV capsid.
 DR InterPro: IPR002521; HCV core.
 DR InterPro: IPR002519; HCV env.
 DR InterPro: IPR002531; HCV NS1.
 DR InterPro: IPR002518; HCV NS2.
 DR InterPro: IPR004109; HCV NS3.
 DR InterPro: IPR000745; HCV NS4a.
 DR InterPro: IPR001490; HCV NS4b.
 DR InterPro: IPR002868; HCV NS5a.
 DR InterPro: IPR002166; HCV Rdrp.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_Psvir.
 DR Pfam: PF01543; HCV capsid; 1.
 DR Pfam: PF01542; HCV core; 1.
 DR Pfam: PF01539; HCV env; 1.
 DR Pfam: PF01560; HCV NS1; 1.
 DR Pfam: PF01538; HCV NS2; 1.
 DR Pfam: PF02907; HCV NS3; 1.
 DR Pfam: PF01006; HCV NS4a; 1.
 DR Pfam: PF01001; HCV NS4b; 1.
 DR Pfam: PF01506; HCV NS5a; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00998; Viral_Rdrp; 1.
 DR ProDom: PD186062; HCV NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR PROSITE: PS05057; RDRP_VIRAL; 1.
 DR PROSITE: PS05021; RDRP_POSITIVE; 1.
 KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
 KW Hydrolyase; Nonstructural protein; Polyprotein;
 KW RNA-directed RNA polymerase; transferase; transmembrane.
 SQ SEQUENCE 3011 AA; 327112 MW; 97E9052C0250463B CRC64;

```

DR DR
DR HSP; P217936; RAGL1; RAGL1B0.1;
DR DR
DR InterPro: IPR001410; DEAD.
DR DR InterPro: IPR002522; HCV.
DR DR InterPro: IPR002519; HCV core.
DR DR InterPro: IPR002511; HCV env.
DR DR InterPro: IPR002531; HCV NS1.
DR DR InterPro: IPR002538; HCV NS2.
DR DR InterPro: IPR004109; HCV NS3.
DR DR InterPro: IPR000745; HCV NS4a.
DR DR InterPro: IPR001430; HCV NS4b.
DR DR InterPro: IPR002868; HCV NS5a.
DR DR InterPro: IPR002166; HCV RDRP.

```

```

DR InterPro: IPR001650; Helicase.C.
DR InterPro: IPR002129; Pyridoxal_deC.
DR InterPro: IPR007099; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01539; HCV_core; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase.C; 1.
DR Pfam: PF00998; Viral_RdRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS00392; DDC_GAD_HDC_YDC; 1.
DR PROSITE: PS05057; RDRP_POSITIVE; 1.
DR PROSITE: PS05021; RDRP_VIRAL; 1.
KW ATP-binding: Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polypotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3015 AA; 328159 MW; B7D23BCIF190663A CRC64;

Query Match 83.7%; Score 851.5; DB 12; Length 3015;
Best Local Similarity 82.4%; Pred. No. 7.7e-74;
Matches 168; Conservative 10; Mismatches 17; Indels 9; Gaps 1;

QY 3 KKGSVVIVGRIN-----LSGDTAYAAQTRGEGCOETSGTRDKNOVEGEVQIVST 53
DB 1009 RGQEILLGPADGMVSKGWRLLAPITAYAAQTRGEGCOETSGTRDKNOVEGEVQIVST 1068

QY 54 ATQTFLATISINGVLWTVYHGAGTRTIASPKGPVTQMTYNDKDLVGVQAPQGSRLTPCT 113
DB 1069 ATQTFLATISINGVLWTVYHGAGTRTIASPKGPVTQMTYNDKDLVGVQAPQGSRLTPCT 1128

QY 114 CGSSDLVLTTRHADVIPVRRGDSRGLSPRPISYLKSGSGGPLLCPCPAGHAVGIFRAAV 173
DB 1129 CGSSDLVLTTRHADVIPVRRGDSRGLSPRPISYLKSGSGGPLLCPCPAGHAVGIFRAAV 1188

QY 174 STRGVAKAVDFIPVESLETTMRSP 197
DB 1189 CTRGVAKAVDFIPVENLGTTRSP 1212

RESULT 9
Q9PMW9 PRELIMINARY; PRT; 3015 AA.
AC Q9PMW9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99420396; PubMed=10489358;
RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;
RT "Hepatitis C virus: an infectious molecular clone of a second major
RT genotype (2a) and lack of viability of intertypic 1a and 2a
RT chimeras."
RL Virology 262:250-263(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Bukh J.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

```

```

CC PROTEIN C AND MNA (BY SIMILARITY).
DR EMBL: AF177039; AAF01181.1; -.
DR EMBL: AF177037; AAF01179.1; -.
DR HSSP: P27958; 1HEI.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRP.
DR InterPro: IPR001650; Helicase.C.
DR InterPro: IPR002129; Pyridoxal_deC.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase.C; 1.
DR Pfam: PF00998; Viral_RdRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS00392; DDC_GAD_HDC_YDC; 1.
DR PROSITE: PS05057; RDRP_POSITIVE; 1.
DR PROSITE: PS05021; RDRP_VIRAL; 1.
KW ATP-binding: Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polypotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3015 AA; 328084 MW; E309F6318067D6CD CRC64;

Query Match 83.7%; Score 851.5; DB 12; Length 3015;
Best Local Similarity 82.4%; Pred. No. 7.7e-74;
Matches 168; Conservative 10; Mismatches 17; Indels 9; Gaps 1;

QY 3 KKGSVVIVGRIN-----LSGDTAYAAQTRGEGCOETSGTRDKNOVEGEVQIVST 53
DB 1009 RGQEILLGPADGMVSKGWRLLAPITAYAAQTRGEGCOETSGTRDKNOVEGEVQIVST 1068

QY 54 ATQTFLATISINGVLWTVYHGAGTRTIASPKGPVTQMTYNDKDLVGVQAPQGSRLTPCT 113
DB 1069 ATQTFLATISINGVLWTVYHGAGTRTIASPKGPVTQMTYNDKDLVGVQAPQGSRLTPCT 1128

QY 114 CGSSDLVLTTRHADVIPVRRGDSRGLSPRPISYLKSGSGGPLLCPCPAGHAVGIFRAAV 173
DB 1129 CGSSDLVLTTRHADVIPVRRGDSRGLSPRPISYLKSGSGGPLLCPCPAGHAVGIFRAAV 1188

QY 174 STRGVAKAVDFIPVESLETTMRSP 197
DB 1189 CTRGVAKAVDFIPVENLGTTRSP 1212

RESULT 10
Q91RR8 PRELIMINARY; PRT; 181 AA.
AC Q91RR8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;

```

```

RN      11
RP      SEQUENCE FROM N.A.
RC      STRAIN-Pt.1Y;
RA      Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT      *Genetic Diversity and response to IFN of the NS3 Protease Gene from
RL      Clinical Strains of the Hepatitis C Virus.*;
RL      Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AF369235; AAK54563.1; -.
DR      InterPro; IPR004109; HCV_NS3.
DR      Pfam; PF02907; HCV_NS3; 1.
KW      Protease.
FT      NON_TER      1      1
FT      NON_TER      181      181
SQ      SEQUENCE      181 AA; 19130 MW; 85D91869299B7C35 CRC64;

Query Match      83.5%; Score 849; DB 12; Length 181;
Best Local Similarity 93.3%; Pred. No. 3.le-75;
Matches 166; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

QY      19 TAYAAQOTRGEEGCOETSQTRDKNQVEGEVQIVSTATQTFATSIINGVLWTVYHGAGTRT 78
DB      4 TAYAAQOTRGLLGCIITSLTGRDKNQVEGEVQIVSTAQTFLATCINGVCWTVYHGAGTRT 63
QY      79 IASPKGPVTQMTYNVDKDLVGMQAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRGDSR 138
DB      64 IASPKGPVTQMTYNVDKDLVGMQAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRGDSR 123
QY      139 GSLLSPRPISYLGSSGGPILCPAGHAGVIFRAAVSTRGVAKAVDFIPVESLETTMRS 196
DB      124 GSLLSPRPISYLGSSGGPILCPAGHAGVIFRAAVSTRGVAKAVDFIPVENLETTMRS 181

RESULT 11
Q91RT5 ID Q91RT5 PRELIMINARY; PRT; 181 AA.
AC Q91RT5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Pt. 4;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT *Genetic Diversity and response to IFN of the NS3 Protease Gene from
RL Clinical Strains of the Hepatitis C Virus.*;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF369218; AAK54543.1; -.
DR InterPro; IPR004109; HCV_NS3.
DR Pfam; PF02907; HCV_NS3; 1.
KW Protease.
FT NON_TER      1      1
FT NON_TER      181      181
SQ SEQUENCE      181 AA; 19130 MW; 85D91869299B7C35 CRC64;

Query Match      83.5%; Score 849; DB 12; Length 181;
Best Local Similarity 93.3%; Pred. No. 3.le-75;
Matches 166; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

QY      19 TAYAAQOTRGEEGCOETSQTRDKNQVEGEVQIVSTATQTFATSIINGVLWTVYHGAGTRT 78
DB      4 TAYAAQOTRGLLGCIITSLTGRDKNQVEGEVQIVSTAQTFLATCINGVCWTVYHGAGTRT 63
QY      79 IASPKGPVTQMTYNVDKDLVGMQAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRGDSR 138
DB      64 IASPKGPVTQMTYNVDKDLVGMQAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRGDSR 123
QY      139 GSLLSPRPISYLGSSGGPILCPAGHAGVIFRAAVSTRGVAKAVDFIPVESLETTMRS 196
DB      124 GSLLSPRPISYLGSSGGPILCPAGHAGVIFRAAVSTRGVAKAVDFIPVENLETTMRS 181

```

```

DB      124 GSLLSPRPISYLGSSGGPILCPAGHAGVIFRAAVSTRGVAKAVDFIPVENLETTMRS 181

RESULT 12
Q91RR5 ID Q91RR5 PRELIMINARY; PRT; 181 AA.
AC Q91RR5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Pt.30;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT *Genetic Diversity and response to IFN of the NS3 Protease Gene from
RL Clinical Strains of the Hepatitis C Virus.*;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF369238; AAK54563.1; -.
DR InterPro; IPR004109; HCV_NS3.
DR Pfam; PF02907; HCV_NS3; 1.
KW Protease.
FT NON_TER      1      1
FT NON_TER      181      181
SQ SEQUENCE      181 AA; 19084 MW; 3B5E8161F2100A72 CRC64;

Query Match      83.3%; Score 847; DB 12; Length 181;
Best Local Similarity 92.7%; Pred. No. 4.9e-75;
Matches 165; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY      19 TAYAAQOTRGEEGCOETSQTRDKNQVEGEVQIVSTATQTFATSIINGVLWTVYHGAGTRT 78
DB      4 TAYAAQOTRGLLGCIITSLTGRDKNQVEGEVQIVSTAQTFLATCINGVCWTVYHGAGTRT 63
QY      79 IASPKGPVTQMTYNVDKDLVGMQAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRGDSR 138
DB      64 IASPKGPVTQMTYNVDKDLVGMQAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRGDSR 123
QY      139 GSLLSPRPISYLGSSGGPILCPAGHAGVIFRAAVSTRGVAKAVDFIPVESLETTMRS 196
DB      124 GSLLSPRPISYLGSSGGPILCPAGHAGVIFRAAVSTRGVAKAVDFIPVESLETTMRS 181

RESULT 13
Q91RR2 ID Q91RR2 PRELIMINARY; PRT; 181 AA.
AC Q91RR2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Pt.4V;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT *Genetic Diversity and response to IFN of the NS3 Protease Gene from
RL Clinical Strains of the Hepatitis C Virus.*;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF369241; AAK54566.1; -.
DR InterPro; IPR004109; HCV_NS3.
DR Pfam; PF02907; HCV_NS3; 1.
KW Protease.
FT NON_TER      1      1
FT NON_TER      181      181
SQ SEQUENCE      181 AA; 19123 MW; 1CAE817345ED809D CRC64;

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 30, 2003, 19:18:33 ; Search time 2560.57 Seconds
(without alignments)
3147.423 Million cell updates/sec

Title: US-09-965-594-18

Perfect score: 1017

Sequence: 1 MKKKGWVIGRINLSGDTA.....YAKAYDFIPVSELTMRSP 197

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-O=Cgn2_1/USPTO.spool/US09965594/runat_29082003_151919_28310/app_query.fasta_1.2872
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOFC=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09965594 -CGN_1_1_14686 -runat_29082003_151919_28310 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*
1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vi:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.mu:*
20: em.om:*
21: em.or:*
22: em.ov:*
23: em.pat:*
24: em.ph:*
25: em.pl:*
26: em.ro:*
27: em.sts:*
28: em.un:*

29: em.vi:*
30: em.htg_hum:*
31: em.htg_inv:*
32: em.htg_other:*
33: em.htg_mus:*
34: em.htg_pin:*
35: em.htg_rod:*
36: em.htg_mam:*
37: em.htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	892.5	87.8	12734	6	ARI179057 Sequence
2	881.5	86.7	1998	6	ARI145264 Sequence
3	878.5	86.4	1998	6	ARI145268 Sequence
4	877.5	86.3	1998	6	ARI145262 Sequence
5	877.5	86.3	1998	6	ARI145263 Sequence
6	874.5	86.0	651	6	ARI145234 Sequence
7	874.5	86.0	1998	6	ARI145266 Sequence
8	874.5	86.0	1998	6	ARI145267 Sequence
9	873.5	85.9	1998	6	ARI145261 Sequence
10	873.5	85.9	2016	6	ARI145269 Sequence
11	872.5	85.8	12734	14	AF268278 Pestivirus
12	871.5	85.7	651	6	ARI145258 Sequence
13	870.5	85.6	651	6	ARI145252 Sequence
14	870.5	85.6	651	6	ARI145253 Sequence
15	870.5	85.6	1998	6	ARI145265 Sequence
16	870.5	85.6	2016	6	ARI145270 Sequence
17	870.5	85.6	648	6	ARI145274 Sequence
18	868	85.3	648	6	ARI145272 Sequence
19	867.5	85.3	651	6	ARI145256 Sequence
20	867.5	85.3	651	6	ARI145257 Sequence
21	867.5	85.3	651	6	ARI145260 Sequence
22	866.5	85.2	651	6	ARI145251 Sequence
23	866	85.2	648	6	ARI145273 Sequence
24	864	85.0	648	6	ARI145271 Sequence
25	863.5	84.9	651	6	ARI145255 Sequence
26	863.5	84.9	651	6	ARI145259 Sequence
27	861	84.7	8157	6	ARI127810 Sequence
28	861	84.7	8157	6	BD081911 Hepatitis
29	859	84.5	1932	6	BD127809 Sequence
30	859	84.5	1932	6	BD081910 Hepatitis
31	858.5	84.4	9646	6	BD110828 Hepatitis
32	858.5	84.4	9646	6	BD069982 Functiona
33	858.5	84.4	9646	14	AF009806 Hepatitis
34	858.5	84.4	12980	6	ARI10831 Sequence
35	858.5	84.4	12980	6	BD069985 Functiona
36	854.5	84.0	5360	6	ARI118686 Sequence
37	854.5	84.0	5360	6	I06434 Sequence 48
38	854.5	84.0	5360	6	I09328 Sequence 8
39	854.5	84.0	5360	6	ARI118692 Sequence
40	854.5	84.0	6785	6	I06440 Sequence 54
41	854.5	84.0	6785	6	I09329 Sequence 10
42	854.5	84.0	7310	6	ARI118696 Sequence
43	854.5	84.0	7310	6	I09331 Sequence 15
44	854.5	84.0	7310	14	HPCPOLYP M32084 Hepatitis C
45	854.5	84.0	8316	6	ARI118703 Sequence

ALIGNMENTS

RESULT 1

ARI179057
LOCUS ARI179057 12734 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6326137.
ACCESSION ARI179057
VERSION ARI179057.1 GI:20220612
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 12734)
AUTHORS Hong, Z., Lai, V.C.H. and Lau, J.Y.N.
TITLE Hepatitis C virus protease-dependent chimeric pestivirus
JOURNAL Patent: US 6326137-A 1 04-DEC-2001;
FEATURES Location/Qualifiers
source
1..12734
/organism="unknown"
BASE COUNT 4032 a 2604 c 3295 g 2803 t
ORIGIN
Alignment Scores:
Pred. No.: 1,11e-65 Length: 12734
Score: 892.50 Matches: 177
Percent Similarity: 92.8% Conservative: 4
Best Local Similarity: 90.7% Mismatches: 11
Query Match: 87.7% Indels: 3
DB: 6 Gaps: 1
US-09-965-594-18 (1-197) x ARI179057 (1-12734)
QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
Db 413 GGTAGTGGTGTATTTGGTGGTAGAATATTTTCTGTGTAGTGTAGTATCATCGGCGCTAC 472
QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnThrSerGlnThrGlyArgAspLys 41
Db 473 GCCAGCAGCAGAGGCGCTCTAGGGGTGAAGATCACCAGCTGACTGCGCGGGACAAA 532
QY 42 AsnGlnValGluGlyValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
Db 533 AACCAAGTGGAGGGTGGAGTCCAGATGCTGTCAACTGCTACCCAAACCTTCTCTGGCAACG 592
QY 62 SerIleAsnGlyValIleUtrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
Db 593 TGCATCAATGGGGTATGCTGGAGCTCTACCCAGGGCGGGAACAGGACCATCGCATCA 652
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrPln 101
Db 653 CCCAAGGGTCTCTCATCCAGATGTATACCAATGTGGACCAAGACCTTGTGGGCTGGCCC 712
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
Db 713 GCTCTCAAGGTTCCCGTCAATGTACACCTGTGACCTGCGGCTCTCGGACCTTTACTG 772
QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyArgGlyAspSerArgGlySerLeu 141
Db 773 GTTACAGACGACCGCAGCTCATTTCCGCTGGCGGCGAGGTATAGCAGGGTAGCCTG 832
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerGlyGlyProLeuLeuCysPro 161
Db 833 CTTTCCCGCGCGCCCATTTCTACCTAACAGGCTCTCGGGGGTCCGCTGTGTGGCCCC 892
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
Db 893 GCGGGACAGCGGTGGGCTATTACAGGCGCGGTGTGCACCCGCTGGAGTGGCCAAAGCG 952
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196
Db 953 GTGGACTTTATCCCTGTGGAGAACCTAGAGAACCAACCATGAGATCC 997
RESULT 2
ARI145264
LOCUS ARI145264 1998 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 105 from patent US 6211338.

ACCESSION ARI45264
VERSION ARI45264.1 GI:15107131
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1998)
AUTHORS Malcolm, B.A., Taremi, S., Shane, J., Weber, P.C. and Yao, N.
TITLE Single-chain recombinant complexes of hepatitis C virus NS3
protease and NS4A cofactor peptide
JOURNAL Patent: US 6211338-A 105 03-APR-2001;
FEATURES Location/Qualifiers
source
1..1998
/organism="unknown"
BASE COUNT 411 a 595 c 569 g 423 t
ORIGIN
Alignment Scores:
Pred. No.: 1.1e-65 Length: 1998
Score: 881.50 Matches: 167
Percent Similarity: 93.37% Conservative: 16
Best Local Similarity: 85.20% Mismatches: 10
Query Match: 86.68% Indels: 3
DB: 6 Gaps: 1
US-09-965-594-18 (1-197) x ARI45264 (1-1998)
QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
Db 64 GGTCTGTGTGTATTTGGTGGTAGAATATTTTCTGTGTAGTGTAGTATCATCGGCGCTAC 123
QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnThrSerGlnThrGlyArgAspLys 41
Db 124 TCCCAACAGACGCGGGCGCTACTTGGTTGCAAGAAGACTAGCCTTACAGCGCGGACNAG 183
QY 42 AsnGlnValGluGlyValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
Db 184 AACCAAGTGGAGGGTGGAGTCCAGTGTGTTCACCGCAACAACTCTCTCTGGCGACC 243
QY 62 SerIleAsnGlyValIleUtrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
Db 244 TCGCTCAACGCGGCTGTGTGGACCTTTACCATGCTGCTGCTCAAGACCTTAGCGCGC 303
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrPln 101
Db 304 CCAAAGGGCGCCATGCCAGATGTACACTAATGTGGACAGGACCTCGTGGCTGGCAG 363
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
Db 364 GCGCCCGCGGGCGCGTCTCTTGCACACCATGACCTGCTGGCAGCTCAGACCTTTACTTG 423
QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyArgGlyAspSerArgGlySerLeu 141
Db 424 GTCACGAGACATGCTGACGTCAATCCGCTGCGCGGCGGCGGACAGTAGGGGAGCCTG 483
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerGlyGlyProLeuLeuCysPro 161
Db 484 CTTCTCCCGCAGCGCTGCTCTACTTTGAAGGCTCTTCCGGTGTGCCACTGCTCGCCT 543
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
Db 544 TCGGGGACGCTGTGGCATCTTCGGGCTGCCGTATGCACCGGGGGGTTTGGCAAGCG 603
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
Db 604 GTGGACTTTGTGCCGCTAGAGTCCATGGAATACTATGCGGCTCTCCG 651
RESULT 3
ARI45268
LOCUS ARI45268 1998 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 109 from patent US 6211338.
ACCESSION ARI45268
VERSION ARI45268.1 GI:15107135

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1998)

AUTHORS Malcolin,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.

TITLE Single-chain recombinant complexes of hepatitis C virus NS3

JOURNAL protease and NS4A cofactor peptide

Patent: US 6211338-A 103 03-APR-2001;

FEATURES Location/Qualifiers

1..1998

BASE COUNT 411 a 595 c 569 g 423 t

ORIGIN /organism="unknown"

Alignment Scores:

Pred. No.: 1.99e-65 Length: 1998

Score: 878.50 Matches: 166

Percent Similarity: 93.37% Conservative: 17

Best Local Similarity: 84.69% Mismatches: 10

Query Match: 86.38% Indels: 3

DB: 6 Gaps: 1

US-09-965-594-18 (1-197) x AR145268 (1-1998)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21

Db 64 GGTCTGTTGTTATTTGTTGTTAGTAATTTATTTATCTGGTAGTAGTATCATCGGCGCTAC 123

QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41

Db 124 TCCCAACAGACGCGGGCCCTACTGTTGTCGAAGAAGACTAGCCTTACAGGCGGGACAAG 183

QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61

Db 184 AACCAAGTCGAGGAGAGGTTTCAGTGGTTTCCACCCACACAAATCTCTCTGGCGACC 243

QY 62 SerIleAsnGlyValLeuThrValThrValThrHisGlyAlaGlyThrArgThrIleAlaSer 81

Db 244 TCGCTCAACGCGGTGTGTGGACCGTTTACCATGTGTGCTGGCTCAAGACCTTAGCGCGC 303

QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101

Db 304 CCNAAAGGGCCCAATCACCAGCATGTACACTAATGTGACACGACCTCTGCGCTGGCAG 363

QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121

Db 364 CGCGCCCGCGGGCGGTCTCTTCCACCATGTCACCTGTGGCAGCTCAGACCTTTACTTG 423

QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141

Db 424 GTCACGAGACATGCTGCTACTTCCGGTGGCGGGCGGCGACAGTAGGGGAGCCTG 483

QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161

Db 484 CTCCTCCCGCAGCGCTGCTCTCTACTTCAAGGGCTCTGCTGGTGTCCACCTCTCGCCCT 543

QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181

Db 544 TCGGGGACGCTGTGGGCATCTCCGGCTGCCCATATGCACCGGGGGGTGGCAAGCG 603

QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197

Db 604 GTGGACTTTGTCCCGTAGAGTCCATGGAAACTACTATGCGGCTCTCCG 651

RESULT 4

AR145262

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

Sequence 103 from patent US 6211338.

AR145262

AR145262

AR145262.1

GI:15107129

Unknown.

1998 bp DNA linear PAT 08-AUG-2001

ORGANISM

Unknown.

Unclassified.

REFERENCE 1 (bases 1 to 1998)

AUTHORS Malcolin,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.

TITLE Single-chain recombinant complexes of hepatitis C virus NS3

JOURNAL protease and NS4A cofactor peptide

Patent: US 6211338-A 103 03-APR-2001;

FEATURES Location/Qualifiers

1..1998

BASE COUNT 410 a 596 c 568 g 424 t

ORIGIN /organism="unknown"

Alignment Scores:

Pred. No.: 2.42e-65 Length: 1998

Score: 877.50 Matches: 167

Percent Similarity: 92.86% Conservative: 15

Best Local Similarity: 85.20% Mismatches: 11

Query Match: 86.28% Indels: 3

DB: 6 Gaps: 1

US-09-965-594-18 (1-197) x AR145262 (1-1998)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21

Db 64 GGTCTGTTGTTATTTGTTGTTAGTAATTTATTTATCTGGTAGTAGTATCATCGGCGCTAC 123

QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41

Db 124 TCCCAACAGACGCGGGCCCTACTTGGTTGCAAGACTACAGCCTTACAGGCGGGACAAG 183

QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61

Db 184 AACCAAGTCGAGGAGAGGTTTCAGTGGTTTCCACCCGACACAAATCTCTCTGGCGACC 243

QY 62 SerIleAsnGlyValLeuThrValThrValThrHisGlyAlaGlyThrArgThrIleAlaSer 81

Db 244 TCGCTCAACGCGGTGTGTGGACCGTTTACCATGTGTGCTGGCTCAAGACCTTAGCGCGC 303

QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101

Db 304 CCNAAAGGGCCCAATCACCAGCATGTACACTAATGTGACACGACCTCTGCGCTGGCAG 363

QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121

Db 364 CGCGCCCGCGGGCGGTCTCTTCCACCATGTCACCTGTGGCAGCTCAGACCTTTACTTG 423

QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141

Db 424 GTCACGAGACATGCTGCTACTTCCGGTGGCGGGCGGCGACAGTAGGGGAGCCTG 483

QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161

Db 484 CTCCTCCCGCAGCGCTGCTCTCTACTTGAAGGGCTCTTCCGGTGTCCACCTCTCTGCCCT 543

QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181

Db 544 TCGGGGACGCTGTGGGCATCTCCGGCTGCCCATATGCACCGGGGGGTGGCAAGCG 603

QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197

Db 604 GTGGACTTTGTCCCGTAGAGTCCATGGAAACTACTATGCGGCTCTCCG 651

RESULT 5

AR145263

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

Sequence 104 from patent US 6211338.

AR145263

AR145263

AR145263.1

GI:15107130

Unknown.

Unknown.

Unclassified.

1998 bp DNA linear PAT 08-AUG-2001

REFERENCE 1 (bases 1 to 1998)
 AUTHORS Malcolim,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.
 TITLE Single-chain recombinant complexes of hepatitis C virus NS3
 protease and NS4A cofactor peptide
 Patent: US 6211338-A 104 03-APR-2001;
 JOURNAL Location/Qualifiers
 FEATURES
 source 1..1998
 Location/Qualifiers
 /organism="unknown"

BASE COUNT 410 a 596 c 568 g 424 t

ORIGIN

Alignment Scores:
 Pred. No.: 2,42e-65 Length: 1998
 Score: 877.50 Matches: 167
 Percent Similarity: 92.86% Conservative: 15
 Best Local Similarity: 85.20% Mismatches: 11
 Query Match: 86.28% Indels: 3
 DB: Gaps: 6

US-09-965-594-18 (1-197) x ARI45263 (1-1998)

Qy 5 GlySerValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
 |||||
 Db 64 GGTCTGTTGTTATTGTTGTTAGTAATATTATCTGTTGTTAGTATACGGCCCTAC 123
 Qy 22 AlaGlnGlnThrArgGlyGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41
 :|||
 Db 124 TCCCAACAGACGCGGGCCTACTTGTGTCATCAAGACTAGCCTTACAGCGCGGACAA 183
 Qy 42 AsnGlnValGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
 |||||
 Db 184 AACAGGTCGAGGAGGTTCCAGGTGTTCCACCGCAACAAATCTCTCTGGCGACC 243
 Qy 62 SerIleAsnGlyValLeuThrPrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
 :|||
 Db 244 TGGCTCAACGCGCTGTTGGACCGTTTACCATTGTTGCTGCTCAAGACCTTAGCGCGC 303
 Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
 |||||
 Db 304 CCAAGGGGCAATACCCAGATGTACACTAATGTGGACACGAGCTCTCGGCTGGCAG 363
 Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
 |||||
 Db 364 CGCGCCCGCGGGCGGTCCTTGCACCATGCACCTGTGGCAGCTCAGACCTTTACTTG 423
 Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
 |||||
 Db 424 GTCACGAGACATGTCACGTCATTCCGGTGCAGCGGGGGGACAGTAGAGGGGAGCCTG 483
 Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
 |||||
 Db 484 CTCTCCCGCAGGCTGTCCTACTTTGAAGGGCTCTTCGGGTGGTCCACTGCTCTGCCCT 543
 Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
 :|||
 Db 544 TCGGGGACGCTGTGGGCATCTTCCGGGTGCGGTATGCACCGGGGGGTTCGGAAGGCG 603
 Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
 |||||
 Db 604 GTGGACTTTGTGCGCGTAGAGTCCATGGAAACTACTATGCGGTCTCCG 651

RESULT 6

ARI45254
 LOCUS 651 bp DNA linear PAT 08-AUG-2001
 DEFINITION Sequence 95 from patent US 6211338.
 ACCESSION ARI45254
 VERSION ARI45254.1 GI:15107121
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.

REFERENCE 1 (bases 1 to 651)
 AUTHORS Malcolim,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.

TITLE Single-chain recombinant complexes of hepatitis C virus NS3
 protease and NS4A cofactor peptide
 Patent: US 6211338-A 95 03-APR-2001;
 JOURNAL Location/Qualifiers
 FEATURES
 source 1..651
 Location/Qualifiers
 /organism="unknown"

BASE COUNT 120 a 187 c 200 g 144 t

ORIGIN

Alignment Scores:
 Pred. No.: 1.18e-65 Length: 651
 Score: 874.50 Matches: 166
 Percent Similarity: 93.33% Conservative: 16
 Best Local Similarity: 85.13% Mismatches: 10
 Query Match: 85.99% Indels: 3
 DB: Gaps: 1

US-09-965-594-18 (1-197) x ARI45254 (1-651)

Qy 5 GlySerValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
 |||||
 Db 64 GGTCTGTTGTTATTGTTGTTAGTAATATTATCTGTTGTTAGTATACGGCCCTAC 123
 Qy 22 AlaGlnGlnThrArgGlyGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41
 :|||
 Db 124 TCCCAACAGACGCGGGCCTACTTGTGTTCAAGAAAGACTAGCCTTACAGCGCGGACAA 183
 Qy 42 AsnGlnValGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
 |||||
 Db 184 AACAGGTCGAGGAGGTTCCAGGTGTTCCACCGCAACAAATCTCTCTGGCGACC 243
 Qy 62 SerIleAsnGlyValLeuThrPrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
 :|||
 Db 244 TGGCTCAACGCGCTGTTGGACCGTTTACCATTGTTGCTGCTCAAGACCTTAGCGCGC 303
 Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
 |||||
 Db 304 CCAAGGGGCAATACCCAGATGTACACTAATGTGGACACGAGCTCTCGGCTGGCAG 363
 Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
 |||||
 Db 364 CGCGCCCGCGGGCGGTCCTTGCACCATGCACCTGTGGCAGCTCAGACCTTTACTTG 423
 Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
 |||||
 Db 424 GTCACGAGACATGTCACGTCATTCCGGTGCAGCGGGGGGACAGTAGAGGGGAGCCTG 483
 Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
 |||||
 Db 484 CTCTCCCGCAGGCTGTCCTACTTTGAAGGGCTCTTCGGGTGGTCCACTGCTCTGCCCT 543
 Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
 :|||
 Db 544 TCGGGGACGCTGTGGGCATCTTCCGGGTGCGGTATGCACCGGGGGGTTCGGAAGGCG 603
 Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196
 |||||
 Db 604 GTGGACTTTGTGCGCGTAGAGTCCATGGAAACTACTATGCGGTCT 648

RESULT 7

ARI45266
 LOCUS 1998 bp DNA linear PAT 08-AUG-2001
 DEFINITION Sequence 107 from patent US 6211338.
 ACCESSION ARI45266
 VERSION ARI45266.1 GI:15107133
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1998)
 AUTHORS Malcolim,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.
 TITLE Single-chain recombinant complexes of hepatitis C virus NS3
 protease and NS4A cofactor peptide

JOURNAL Patent: US 6211338-A 107 03-APR-2001;

FEATURES

source Location/Qualifiers

BASE COUNT 410 a 596 c 568 g 424 t
ORIGIN 1..1998 /organism="unknown"

Alignment Scores: 4.35e-65 Length: 1998
Pred. No.: 874.50 Matches: 166
Score: 92.86% Conservative: 16
Percent Similarity: 84.69% Mismatches: 11
Best Local Similarity: 85.99% Indels: 3
Query Match: 6 Gaps: 1
DB: 1

US-09-965-594-18 (1-197) x ARI45266 (1-1998)

Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThralaTyr 21
Db 64 GGTCTGTTGTTATTGTTGTTAGATTATTTATCTGGTAGTGTAGTATCATCGGCTTAC 123
Qy 22 AlaGlnGlnThrArgGlyGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41
Db 124 TCCCAACACAGCGGGGCGCTTACTTGGTTGCAAGATCATAGCTTTACAGCGGGGACAA 183
Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
Db 184 AACCGGTGAGGAGAGGTTCCAGGTGGTTCCACCGCAACACATCTCTCTGGCGACC 243
Qy 62 SerIleAsnGlyValLeuThrPrhrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
Db 244 TGGCTCAACGGCGTGTGGACCGTTTACCATTGGTGGCTCAAGACCTTACCGCGC 303
Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101
Db 304 CCAAGGGGCAATCATCCAGATGTACACTATGTGGACAGGACCTCGTGGCTGGCAG 363
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
Db 364 GCGCCCGCGGGCGGCTTCTTGACACCATGCACCTGTGGCAGCTCAGACCTTTACTTG 423
Qy 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141
Db 424 GTCCAGAGACATCTCAGCTCATTCGGTGGCGGGCGGCGAGAGAGGAGCGCTG 483
Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
Db 484 CTCTCCCGCAGGCGCTGCTCTACTTGAAGGGCTCTGCTGGTGGTCCACTGCTCTGCCCT 543
Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
Db 544 TCGGGGACGCTGTGGCATCTTCCGGGCTGCGGTATGACCCCGGGGGGTTCGGAAGCG 603
Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
Db 604 GTGGACTTGTGCGCGTAGAGTCCATGGAACTACTATCGGCTCTCCG 651

RESULT 8
ARI45267
LOCUS ARI45267 1998 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 108 from patent US 6211338.
ACCESSION ARI45267
VERSION ARI45267.1 GI:15107134
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1998)
AUTHORS Malcolin B.A., Taremi S.Shane., Weber, P.C. and Yao, N.
TITLE Single-chain recombinant complexes of hepatitis C virus NS3 protease and NS4A cofactor peptide
JOURNAL Patent: US 6211338-A 108 03-APR-2001;
FEATURES Location/Qualifiers

source 1..1998 /organism="unknown"

BASE COUNT 410 a 596 c 568 g 424 t
ORIGIN 1..1998

Alignment Scores: 4.35e-65 Length: 1998
Pred. No.: 874.50 Matches: 166
Score: 92.86% Conservative: 16
Percent Similarity: 84.69% Mismatches: 11
Best Local Similarity: 85.99% Indels: 3
Query Match: 6 Gaps: 1
DB: 1

US-09-965-594-18 (1-197) x ARI45267 (1-1998)

Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThralaTyr 21
Db 64 GGTCTGTTGTTATTGTTGTTAGATTATTTATCTGGTAGTGTAGTATCATCGGCTTAC 123
Qy 22 AlaGlnGlnThrArgGlyGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41
Db 124 TCCCAACACAGCGGGGCGCTTACTTGGTTGCAAGATCATAGCTTTACAGCGGGGACAA 183
Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
Db 184 AACCGGTGAGGAGAGGTTCCAGGTGGTTCCACCGCAACACATCTCTCTGGCGACC 243
Qy 62 SerIleAsnGlyValLeuThrPrhrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
Db 244 TGGCTCAACGGCGTGTGGACCGTTTACCATTGGTGGCTCAAGACCTTACCGCGC 303
Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101
Db 304 CCAAGGGGCAATCATCCAGATGTACACTATGTGGACAGGACCTCGTGGCTGGCAG 363
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
Db 364 GCGCCCGCGGGCGGCTTCTTGACACCATGCACCTGTGGCAGCTCAGACCTTTACTTG 423
Qy 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141
Db 424 GTCCAGAGACATCTCAGCTCATTCGGTGGCGGGCGGCGAGAGAGGAGCGCTG 483
Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
Db 484 CTCTCCCGCAGGCGCTGCTCTACTTGAAGGGCTCTGCTGGTGGTCCACTGCTCTGCCCT 543
Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
Db 544 TCGGGGACGCTGTGGCATCTTCCGGGCTGCGGTATGACCCCGGGGGGTTCGGAAGCG 603
Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
Db 604 GTGGACTTGTGCGCGTAGAGTCCATGGAACTACTATCGGCTCTCCG 651

RESULT 9
ARI45261
LOCUS ARI45261 1998 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 102 from patent US 6211338.
ACCESSION ARI45261
VERSION ARI45261.1 GI:15107128
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1998)
AUTHORS Malcolin B.A., Taremi S.Shane., Weber, P.C. and Yao, N.
TITLE Single-chain recombinant complexes of hepatitis C virus NS3 protease and NS4A cofactor peptide
JOURNAL Patent: US 6211338-A 102 03-APR-2001;
FEATURES Location/Qualifiers
source 1..1998 /organism="unknown"

BASE COUNT 409 a 597 c 567 g 425 t
ORIGIN

Alignment Scores:

Pred. No.: 5,29e-65 Length: 1998
Score: 873.50 Matches: 167
Percent Similarity: 92.35% Conservative: 14
Best Local Similarity: 85.20% Mismatches: 12
Query Match: 85.89% Indels: 3
DB: 6 Gaps: 1

US-09-965-594-18 (1-197) x AR145261 (1-1998)

QY 5 GlySerValValIleValGlyArgGlyLeuSerGlyAsp-----ThrAlaTyr 21
DB 64 GGTCTGTTGTTATGTTGGTAGAATATTATTCGTAGTGGTAGTATCATCGGCTAC 123
QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41
DB 124 TCCCAACAGACGCGGGGCTACTTGGTTGCATCATCATCTAGCCTTACAGCGCGGACAAG 183
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
DB 184 AACCAAGTTCGAGGAGAGGTTTCAGGTGGTTTCCACCGCAACACAAATCCTTCCCTGGGGACC 243
QY 62 SerIleAsnGlyValLeuTyrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
DB 244 TCGGTCAACGCGCGTGTGTGACCGTTTACCNTGCTGGCTCAAGACCTTAGCGCGC 303
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101
DB 304 CCAAGGGGCGCAATCACCAGATACACTAATGTGGACCAAGACCTTCGTCGGCTGGCAG 363
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
DB 364 GCGCCCGCGGGCGCGTCTTGCACCATGCACTGTGCACCTGCGACCTCAGACCTTACTTG 423
QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
DB 424 GTACGAGACATGCTGACGTCTTCCGTCGCGCGCGCGCGACAGTAGGGGAGCGTG 483
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerGlyGlyProLeuLeuCysPro 161
DB 484 CTTCTCCCGACGCGCTCTCTACTTGAAGGCTCTTCGGGTGCTGCCTCTGCCCT 543
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
DB 544 TCGGGGACGCTGTGGCATCTTCGGGCTCGCGTATGCACCGCGGGGGTTGCGAAGCG 603
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
DB 604 GTGGACTTTGTCCCGTAGAGTCCATGTAAGAACTACTATGCGGTCTCCG 651

RESULT 10
LOCUS AR145269 2016 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 110 from patent US 6211338.
ACCESSION AR145269
VERSION AR145269.1 GI:15107136
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.

REFERENCE
1 (bases 1 to 2016)
AUTHORS Malcolim.B.A., Taremi.S.Shane., Weber.P.C. and Yao.N.
TITLE Single-chain recombinant complexes of hepatitis C virus NS3
protease and NS4A cofactor peptide
JOURNAL Patent: US 6211338-A 110 03-APR-2001;
FEATURES Location/Qualifiers
1..2016 /organism="unknown"
source 412 a 603 c 570 g 431 t

BASE COUNT
ORIGIN

Alignment Scores:

Pred. No.: 5,35e-65 Length: 2016
Score: 873.50 Matches: 167
Percent Similarity: 92.35% Conservative: 14
Best Local Similarity: 85.20% Mismatches: 12
Query Match: 85.89% Indels: 3
DB: 6 Gaps: 1

US-09-965-594-18 (1-197) x AR145269 (1-2016)

QY 5 GlySerValValIleValGlyArgGlyLeuSerGlyAsp-----ThrAlaTyr 21
DB 82 GGTCTGTTGTTATGTTGGTAGAATATTATTCGTAGTGGTAGTATCATCGGCTAC 141
QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41
DB 142 TCCCAACAGACGCGGGGCTACTTGGTTGCATCATCATCTAGCCTTACAGCGCGGACAAG 201
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
DB 202 AACCAAGTTCGAGGAGAGGTTTCAGGTGGTTTCCACCGCAACACAAATCCTTCCCTGGGGACC 261
QY 62 SerIleAsnGlyValLeuTyrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
DB 262 TCGGTCAACGCGCGTGTGTGACCGTTTACCATGCTGGCTCAAGACCTTAGCGCGC 321
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101
DB 322 CCAAGGGGCGCAATCACCAGATGACACTAATGTGGACCAAGACCTTCGTCGGCTGGCAG 381
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
DB 382 GCGCCCGCGGGCGCGTCTTGCACCATGCACTGTGCACCTCAGACCTTACTTG 441
QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
DB 442 GTACGAGACATGCTGACGTCTTCCGTCGCGCGCGCGCGACAGTAGGGGAGCGTG 501
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerGlyGlyProLeuLeuCysPro 161
DB 502 CTTCTCCCGACGCGCTCTCTACTTGAAGGCTCTTCGGGTGCTGCCACTGCTGCCCT 561
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
DB 562 TCGGGGACGCTGTGGCATCTTCGGGCTCGCGTATGCACCGCGGGGGTTGCGAAGCG 621
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
DB 622 GTGGACTTTGTCCCGTAGAGTCCATGTAAGAACTACTATGCGGTCTCCG 669

RESULT 11
LOCUS AF268278 12734 bp RNA linear VRL 12-JUL-2000
DEFINITION Pestivirus type 1, complete genome.
ACCESSION AF268278
VERSION AF268278.1 GI:9049956
KEYWORDS
SOURCE Pestivirus type 1
ORGANISM Pestivirus type 1
VIRUSES: ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Pestivirus.

REFERENCE
1 (bases 1 to 12734)
AUTHORS Lai,V.C., Zhong,W., Skelton,A., Ingravallo,P., Vassiliev,V.,
Donis,R.O., Hong,Z. and Lau,J.Y.
TITLE Generation and characterization of a hepatitis C virus NS3
protease-dependent bovine viral diarrhoea virus
JOURNAL J. Virol. 74 (14), 6339-6347 (2000)
MEDLINE 20323484
PUBMED 10864644
REFERENCE 2 (bases 1 to 12734)
AUTHORS Lai,V.C.H. and Hong,Z.
TITLE Direct Submission

JOURNAL Submitted (16-MAY-2000) Antiviral Therapy, Schering-Plough Research Institute, 2015 Galloping Hill Road, Kenilworth, NJ 07033-0539, USA

FEATURES
source

```

1..12734
/organism="Pestivirus type 1"
/mol_type="genomic RNA"
/db_xref="taxon:11099"
1..385
386..12508
/codon_start=1
/product="polyprotein"
/protein_id="AAF92566.1"
/db_xref="gi:9049957"
/translation="MELNTNMGSGSVVIGRIVLVSGSGSITACAQOTRGLGCGKITSL
TGROKNQVEGVQIVSTAVTQOTFLATCINGVCVTVYHAGTRITIASPRGPVQIOMYTNL
QDGLWMPAQSGRSLLPTCGSSDLVITRHHANVIPVRRGDSRGLSLSPRSYLYLKG
SSGGLPCPAGHAYLPCFACTGTRGAVKADVEIPVENLETTTRSGSGADTEDVCCSM
SYSDTKPEGATKKTQKPDRLERCMMKIVPKESEKSKTKPPDATIVBEVKYQVPRKK
GKTAKNTQDGLYNNKKNPQBSRKKLEKALLANAIILVLFQVMTGENTIQWNLQDGR
TEGIGRAFQGVYNNKSHGIMPEKICGVSPSHLALHAIILVLEQVMTGENTIQWNLQD
LORHEMKNHGKNYNIETPIILVNNRTOANLETQGPPECAVTCRYDRASDLNVVYTG
RDSPTPLTKCKKKGNFPAFTLMRGPNFTIAASDLFKHEHRIISFQDFTLYLVYDGL
TNSLGEARQGTAKITLVGKQGLTGKGLKENKMTGFGAVASPYCDVDPRKIGYIWT
KNCYPACLPKNTKITLVGKQGLTNADEACKILHEMGHLSVEALLSLVLSDFAPETASV
MILLHFSIPQSHVDVMDCKTQNLNIVELTADVIPGSVNNLGNKWCIPNNWMPVET
TVLLAPEYSQVAKVLRAIRDLTRIINNAATTATVILCLVLYRGQWVQYILHLLIT
GVQGHLDCKPFTAIANDERIGQAGELUPTIMKESYSGKLEKIDVIAQEDCGKLM
YLQRCETRIYALIHLTTRQYVFKKLFEDRKQEDVEMDNFNEFKGLPCDAQKPIV
RGKENTILLNGPAOTVCPYIGMTQVTSCTSFNMDDLATTVYTRYRSPKPCDQKPIV
ORNGEDLHNGILCQNVPCVGDQLLYKGGSEISCKWCYQGFKESEGLPHYPICGKCL
ENYELVUDTSCNREGVATVPQGLTKCKTIGTKTQVIAADTKLQMPMRPEYILSS
EGPVSKTACTNYTKLANKYFEPDRSFQOYMLKGEVYWFDELVTDDHRRDYFAESI
LVVYLLHRESYKKNWLLYHILVPIKSVYILLMIGDVYKASGGQGYLCKLIDLCFT
TVYLVIGLIIARSDTPIVPLTINAAIRVTELTQPGVDIAVAMTITLLASTSYTVR
YFRKMLQCLISLVSGVFLYLSIYIGRIEMPEYIPNNRPDLTLLLYLLISTYTVR
WKVDVAGLLUQCVPIILLVTLTMADFTLTLILPTLYBELVLYLKTVRTDIERSMLGQ
IDYTRVDSIYDVSDEGEVYFPPSRQAQGNFSILLPLIKATILSCVSSKKWQIYMSY
LTLQPMYPMHRKVTYSESGGNNIISRLVAALIELNWSEEEESKGLKFKYLLSLGRNL
LIIKHVNEVYKSWGEEVYGNPKITIKASTLSKSRHCLICVCEGRKNGTC
PKCGRHGXPTCGMSLADFEERHYKRIPIREGNEGCMSCRCQKRRFEMDRPKSAR
YCAECNRHPAEBGDFAESSMGLKITYPALMDGKVYDITEMAGCQVGSIGDTHRV
TSIFSFRMGPPROEYNGFVQYARGOLFTRPLVATKYVMLMVMGNLGEISDNLHSL
GWIILRGPAVCKITIEHEKCHILIDLTAFGIMPRGTTIPRAPVPTSLIKYVRGLE
TGWAYTHOGGISYVDHVTAGKDLVDSMGRTVVCCNNRRLDTEYGVKTDSCGDP
GARCYLPAEAVNTSGSAGVHLQKGTGECTVTSAGTAPFDLKLKNGKSGQLPIFLA
ASSGRVGRVYKGNKESSKPKIMSGIQTYSKNTADTEMVKITISNRNGDKFOITLPE
TGACTHTLPKAVIEIGRHRFVLIPIRAAEVSYQYMLRKHPSISPMIGDKME
GDMATGITYASGIFCQNPQPKLAARVAYEYIFELDHCHCATPEOLAIKRIHFSSE
IRVAMTATPAGSVTQKRAPEFTAPEYKMGDLGSLFDLTAGUKIPDMCKGNM
IRPDRNMVAYEAKKLAAGKNGSYGIEDPANLRVVTQSQYPIVATNAIESGVT
LPFDLTVTDGLCEKRVRSKSPFIYVTLGRMAVTVGEQARGRVGRVYKRGYR
SQETAPKQGYHDLLQAGVQIEDGNVTKSFREHNTDMSLEDESSLITQLEIANN
LLISBDLPAAVNTMARTDREPITQLAANSYEQVVPLEPKIRNGEYDITQYNSFLN
AKPLGEDVPVYIATEDDELADVLIGDDPDGQGVVETCKALKQVTGLSAEANLL
VALFCYCYQALSRHPRMIDITYTIEDQRLEDTHLQYAPNAIKQOTGTETSELKAS
GOVEKIMGASIDYANGGLEFVKSQNEKIKTAPLFRKNAEAKGVQKFDLSLENKEE
IIRYGLWGTHTALYKISARLGHETATILVYLWAFGSGSVSDHVPQAAVLYPYVYV
MKNKPSFGDSTQGEGRRAFSVTSALATYTKIYNNHLSRVKVEPAAVLYPYATSA
LKMFTPTRESLVSTITTYKTYLSIRKKSGLDGTGICSAEMELSQNPQSVGIVSM
LCVGAIAAHNAIESSEOKTILMAKVFYNNLDQAAIDELVKNENKPIIMAFEAQVOTI
GNPLRIYHLXGYVYKGEAKELSERVAGNLFLLIMFEAFELIGMDSQCKIINLSON
YILDLIYGLHKIINRGLKKNVLCNAPAPFSCDPTFSBERTFLPDNLNLRVETRCPGY
EMKAKIRNVGGLIKTYIEESGPELGNARPPSCMTWPSDERITFLPDNLNLRVETRCPGY
HYTKGVATRIDYSKGLMILLATDKWEVEHGVITRLAKRTYGVGFNGAYLGDPEHNRAL
ERDCATITKNTVOFLMKKGAFTYDITISNLRTILVLRHNNLEEREIEPTATVTVML
AYTFYNEVDGVTIKPVLGRPIYDPDVQINIQPEQVQDTSVEGVTIGIRETLMTTQVTP
VLEKVEPDASDNQNSVYIGLOEYNGPGCIQIOTHTLTREIHNRDARPPIMVSGNSRIS
NRAKTRARNILYTGNDNREIBDLMAAGRLVALRDYDPELSEWDFKGTFLDREALE
ALSLQCPKQVYTDVARNLIEQKQDVIEPNWFASDDPVELEVALANDRITFLVDGVE
VKDQAKALGATQDTRIIKEVSGRISYAMKLSWFLQASNNQMSUPLPEELLHRCPPAT
KSNKGHASATYQLAQGNMEPLCGGVHLGTTIPARRVKIPHYEAYLKLDFTEEEKXP
VKDVTVIREHNNKWLKIRFOCNLNTKMLNPLGRSEOLDREGRKRNINHQIGTINSS

```


JOURNAL protease and NS4A cofactor peptide
Patent: US 6211338-A 94 03-APR-2001;

FEATURES

Location/Qualifiers
1. .651

BASE COUNT 119 a 188 c 199 g 145 t

ORIGIN /organism="unknown"

Alignment Scores:

Pred. No.: 2.58e-65 Length: 651
Score: 870.50 Matches: 166
Percent Similarity: 92.82% Conservative: 15
Best Local Similarity: 85.13% Mismatches: 11
Query Match: 85.59% Indels: 3
DB: 6 Gaps: 1

US-09-965-594-18 (1-197) x AR145253 (1-651)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
DB 64 GGTTCGTGTTATTTGTTAGTAATTTATTTCTGTAGTAGTATACAGGCGCTAC 123
QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41
DB 124 TCCACAGAGCGCGGCGCTACTTGGTTCATCAAGACTAGCCTTACAGGCGCGGACAAG 183
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
DB 184 AACCAAGTTCGAGGAGAGGTTCCAGTGTTCACCGCAACAACTCTTCCTGGCGACC 243
QY 62 SerIleAsnGlyValLeuThrValThrValThrHisGlyAlaGlyThrArgThrIleAlaSer 81
DB 244 TCGCTCAACGCGGTGTTGGACCGTTTACCATGCTGGCTGCTCAAGACCTTAGCGCGC 303
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101
DB 304 CCAAGGGGCCAATCACCAGATGTACACTAATGTGGACCAAGACCTCGTCGCGTGGCAG 363
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
DB 364 GCGGCGCGCGCGCGCTTCCTTGACACCATGCACTGTGCGAGCTCAGACCTTTACTTG 423
QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
DB 424 GTCACGACATGCTGACGTCTTCGCGTGGCGCGCGGCGGACAGTAGGGGAGCGCTG 483
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
DB 484 CTCCTCCCGCGCGCTGCTCTACTTGAAGGGCTCTTCGGGTGGTCCACTGCTCGCCCT 543
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
DB 544 TCGGGGCGAGCTGTGGGCACTCTTCGGGCTGCCGTATGCACCGGGGGGTTCGGAAGCG 603
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196
DB 604 GTGGACTTGTGCGGTAGATCCATGGAACTACTATGCGGTCT 648

RESULT 15

AR145265

LOCUS

DEFINITION

Sequence 106 from patent US 6211338.

ACCESSION

AR145265

VERSION

AR145265.1

KEYWORDS

Unknown.

SOURCE

Unknown.

ORGANISM

Unclassified.

REFERENCE

1 (bases 1 to 1998)

AUTHORS

Malcolm, B.A., Taremi, S., Shane, Weber, P.C. and Yao, N.

TITLE

Single-chain recombinant complexes of hepatitis C virus NS3

protease and NS4A cofactor peptide

Patent: US 6211338-A 106 03-APR-2001;

JOURNAL

FEATURES Location/Qualifiers
Source 1. .1998

BASE COUNT 409 a 597 c 567 g 425 t

ORIGIN

Alignment Scores:
Pred. No.: 9.53e-65 Length: 1998
Score: 870.50 Matches: 166
Percent Similarity: 92.35% Conservative: 15
Best Local Similarity: 84.69% Mismatches: 12
Query Match: 85.59% Indels: 3
DB: 6 Gaps: 1

US-09-965-594-18 (1-197) x AR145265 (1-1998)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
DB 64 GGTTCGTGTTATTTGTTAGTAATTTATTTCTGTAGTAGTATACAGGCGCTAC 123
QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41
DB 124 TCCACAGAGCGCGGCGCTACTTGGTTCATCACTACCTTACAGGCGCGGACAAG 183
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
DB 184 AACCAAGTTCGAGGAGAGGTTCCAGTGTTCACCGCAACAACTCTTCCTGGCGACC 243
QY 62 SerIleAsnGlyValLeuThrValThrValThrHisGlyAlaGlyThrArgThrIleAlaSer 81
DB 244 TCGCTCAACGCGGTGTTGGACCGTTTACCATGCTGGCTGCTCAAGACCTTAGCGCGC 303
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101
DB 304 CCAAGGGGCCAATCACCAGATGTACACTAATGTGGACCAAGACCTCGTCGCGTGGCAG 363
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
DB 364 GCGGCGCGCGCGCGCTTCCTTGACACCATGCACTGTGCGAGCTCAGACCTTTACTTG 423
QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
DB 424 GTCACGACATGCTGACGTCTTCGCGTGGCGCGCGGCGGACAGTAGGGGAGCGCTG 483
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
DB 484 CTCCTCCCGCGCGCTGCTCTACTTGAAGGGCTCTTCGGGTGGTCCACTGCTCGCCCT 543
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
DB 544 TCGGGGCGAGCTGTGGGCACTCTTCGGGCTGCCGTATGCACCGGGGGGTTCGGAAGCG 603
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
DB 604 GTGGACTTGTGCGGTAGATCCATGGAACTACTATGCGGTCTCGG 651

Search completed: August 31, 2003, 00:46:23

Job time : 2569.57 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 30, 2003, 19:13:57 ; Search time 182.939 Seconds
(without alignments)
2906.924 Million cell updates/sec
Title: US-09-965-594-18
Perfect score: 1017
Sequence: 1 MKKGSVVIVGRINLSGDTA.....VAKAVDFIPVESLETHRSP 197

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-O=/cgn2.1/USPTO_pool/US09965594/runat_29082003.151918_28302/app_query.fasta_1.2872
-DB=N_Geneseq_19Jun03 -QWTF=fastp -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09965594 -RCGN_1_1412.erunat_29082003.151918_28302 -NCPU=6 -ICPU=3
-NO_MMAB -LARGQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_19Jun03:
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1017	100.0	594	21	AAA73332	Hepatitis C virus
2	1005	98.8	594	21	AAA73333	Hepatitis C virus
3	1002	98.5	594	21	AAA73331	Hepatitis C virus
4	995	97.8	594	21	AAA73334	Hepatitis C virus
5	985	96.9	594	21	AAA73330	Hepatitis C virus
6	951	93.5	588	21	AAA73329	Hepatitis C virus
7	946	93.0	594	21	AAA73335	Hepatitis C virus
8	912	89.7	588	21	AAA73328	Hepatitis C virus
9	892.5	87.8	12734	24	ABA95615	Chimeric BVDV/HCV
10	881.5	86.7	1998	20	AAH80355	HCV NS4A-NS3 compl
11	878.5	86.4	1998	20	AAH80359	HCV NS4A-NS3 compl
12	877.5	86.3	1998	20	AAH80353	HCV NS4A-NS3 compl
13	877.5	86.3	1998	20	AAH80354	HCV NS4A-NS3 compl
14	874.5	86.0	612	25	ABX15706	Anti-viral synthet
15	874.5	86.0	651	20	AAH80345	HCV NS4A-NS3 compl
16	874.5	86.0	1998	20	AAH80357	HCV NS4A-NS3 compl
17	874.5	86.0	1998	20	AAH80358	HCV NS4A-NS3 compl
18	873.5	85.9	1998	20	AAH80352	HCV NS4A-NS3 compl
19	873.5	85.9	2013	20	AAH80360	HCV NS4A-NS3 compl
20	871.5	85.7	651	20	AAH80349	HCV NS4A-NS3 compl
21	870.5	85.6	651	20	AAH80343	HCV NS4A-NS3 compl
22	870.5	85.6	651	20	AAH80344	HCV NS4A-NS3 compl
23	870.5	85.6	1998	20	AAH80356	HCV NS4A-NS3 compl
24	870.5	85.6	2016	20	AAH80361	HCV NS4A-NS3 compl
25	870	85.5	648	20	AAH80365	HCV NS4A-NS3 compl
26	868	85.3	648	20	AAH80363	HCV NS4A-NS3 compl
27	867.5	85.3	650	20	AAH80347	HCV NS4A-NS3 compl
28	867.5	85.3	651	20	AAH80348	HCV NS4A-NS3 compl
29	867.5	85.3	651	20	AAH80351	HCV NS4A-NS3 compl
30	866.5	85.2	651	20	AAH80342	HCV NS4A-NS3 compl
31	864	84.9	648	20	AAH80362	HCV NS4A-NS3 compl
32	863.5	84.9	650	20	AAH80346	HCV NS4A-NS3 compl
33	863.5	84.9	651	20	AAH80350	HCV NS4A-NS3 compl
34	861	84.7	8145	20	AAH23259	Plasmid pET-BS(+)/
35	859	84.5	1933	20	AAH23258	HCV NS3 DNA. Hepa
36	858.5	84.4	9646	19	AAV59361	Hepatitis C virus
37	858.5	84.4	9646	24	ABK87285	cdNA encoding hepa
38	858.5	84.4	12980	19	AAV59364	Hepatitis C virus
39	858.5	84.4	12980	24	ABK87286	Hepatitis C virus
40	858.5	84.4	16622	21	AAZ36212	Nucleotide sequenc
41	854.5	84.0	5300	10	AAH92097	Combined open read
42	854.5	84.0	5360	10	AAH90327	Hepatitis C virus
43	854.5	84.0	6905	10	AAH92103	Combined open read
44	854.5	84.0	7310	10	AAH92106	Combined open read
45	854.5	84.0	7310	10	AAH90336	Composite hepatiti

ALIGNMENTS

RESULT 1
AAA73332
ID AAA73332 standard; DNA; 594 BP.
XX
AC AAA73332;
XX
XX
DT 19-DEC-2000 (first entry)
XX
DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #5.
XX
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW liver failure; liver cancer; mutant; mutein; ds.
XX
OS Hepatitis C virus.
OS Synthetic.
FH Key Location/Qualifiers

```

FT CDS 1..594
PT /*tag= a
XX /product= "NS4A-NS3 fusion protein #5"
XX WO2000040707-A1.
XX
XX 13-JUL-2000.
XX
XX 06-JAN-2000: 2000WO-US00345.
XX
XX 08-JAN-1999: 99US-0115271.
XX
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX WPI: 2000-465976/40.
XX P-PSDB: AAB15223.
XX
XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
XX substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
XX amino acid, useful for screening inhibitors that may treat hepatitis C
XX
XX Claim 26; Fig 15; 66pp; English.
XX
XX The present sequence is the coding sequence for a mutated version of a
XX fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
XX protease enzymes. These proteins are both essential for the replication
XX of the virus, acting to cleave its replicative proteins from the
XX polyprotein produced from the HCV genome. Inhibitors of the two proteins
XX should be effective as antiviral treatments of HCV infection. This is
XX useful as HCV can lead to chronic liver disease such as cirrhosis, liver
XX failure and liver cancer. The present invention concerns a number of NS3
XX mutants and NS3-NS4A fusion proteins which can be used to identify
XX inhibitors of this type, as well as enabling structural studies of the
XX protease and protease-inhibitor complexes. The protein produced from this
XX sequence contains the alpha-helix0-1 variant.
XX
XX Sequence 594 BP: 105 A; 189 C; 153 G; 147 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 1 34e-86 Length: 594
XX Score: 1017.00 Matches: 197
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 100.00% Indels: 0
XX DB: 21 Gaps: 0
XX
XX US-09-965-594-18 (1-197) x AAA73332 (1-594)
XX
XX 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
XX 1 ATGAAAAAAGGATCCGTTGTTATCGTCGCCGCTATCAACCTGCGGGTACACCGCT 60
XX
XX 21 TyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlnThrSerGlnThrGlyArgAsp 40
XX 61 TAGCTCAGCAGACTCGAGGTGAGGAGGTGCCAAGAAACCTCCAGACCGCTCGTGAC 120
XX
XX 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
XX 121 AAAAACCCAGGTTCAGGTGAAGTTTCAGATCGTTTCCACCGCTACCCAGACCTTCCTGGCT 180
XX
XX 61 ThrSerIleAsnGlyValLeuThrThrValThrValThrHisGlyAlaGlyThrArgThrIleAla 80
XX 181 ACCTCCATCAACCGGTGTTCTGTGGACCGTTTACCACGGTGGTGGTACCGTACCATCGCT 240
XX
XX 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr 100
XX 241 TCCCGAAGGTCGGTTACCCAGATGTACACCAAGCTTGACAAAGACCTGGTGGTGG 300
XX
XX 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
XX

```

```

Db 301 CAGGCTCCCGAGGGTTCCTCCGTCCGCTGACCCCGTGCACCTGGGGTTCCTCGACCTGTAC 360
QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
Db 361 CTGGTTACCGCTCAGCGTGACGTTATCCCGGTTCGTCGTCGGTGACTCCCGGTGGTTCC 420
QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
Db 421 CTGCTGTCCCGCGTCGATCTCCCTACCTGAAGGTTCCTCCGGTGGTCCGCTGTGGC 480
QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaIleValSerThrArgGlyValAlaLys 180
Db 481 CCGGCTGGTCACGCTGTTGGTATCTTCGTCGTCTGCTTCCACCCGTTGGTGTCTAA 540
QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
Db 541 GCTGTTGACTTCATCCCGGTTCGAACTCCCTGGAACCCACCATCGTTCCTCCCG 591
XX
XX RESULT 2
XX AAA73333
XX ID AAA73333 standard; DNA; 594 BP.
XX AC AAA73333;
XX
XX 19-DEC-2000 (first entry)
XX
XX Hepatitis C virus NS4A-NS3 fusion protease coding sequence #6.
XX
XX Hepatitis; NS3 protease; viral replication; chronic liver disease;
XX liver failure; liver cancer; mutant; mutain; ds.
XX
XX Hepatitis C virus.
XX OS Synthetic.
XX
XX FH Key Location/Qualifiers
XX CDS 1..594
XX FT /*tag= a
XX FT /product= "NS4A-NS3 fusion protein #6"
XX
XX WO2000040707-A1.
XX
XX 13-JUL-2000.
XX
XX 06-JAN-2000: 2000WO-US00345.
XX
XX 08-JAN-1999: 99US-0115271.
XX
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX WPI: 2000-465976/40.
XX DR P-PSDB: AAB15224.
XX
XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
XX substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
XX amino acid, useful for screening inhibitors that may treat hepatitis C
XX
XX Claim 26; Fig 16; 66pp; English.
XX
XX The present sequence is the coding sequence for a mutated version of a
XX fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
XX protease enzymes. These proteins are both essential for the replication
XX of the virus, acting to cleave its replicative proteins from the
XX polyprotein produced from the HCV genome. Inhibitors of the two proteins
XX should be effective as antiviral treatments of HCV infection. This is
XX useful as HCV can lead to chronic liver disease such as cirrhosis, liver
XX failure and liver cancer. The present invention concerns a number of NS3
XX mutants and NS3-NS4A fusion proteins which can be used to identify
XX inhibitors of this type, as well as enabling structural studies of the
XX protease and protease-inhibitor complexes. The protein produced from this
XX sequence contains the alpha-helix0-7 variant.
XX

```


XX SQ Sequence 594 BP; 104 A; 101 C; 152 G; 147 T; 0 other;

Alignment Scores:
 Pred. No.: 1.8e-85 Length: 594
 Score: 1005.00 Matches: 194
 Percent Similarity: 99.49% Conservativity: 2
 Best Local Similarity: 98.48% Mismatches: 1
 Query Match: 98.82% Indels: 0
 DB: 21 Gaps: 0

US-09-965-594-18 (1-197) x AAA73333 (1-594)

QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
 DB 1 ATGAAAAAAGAGTCCGTTGTTATCGCGCGTATCAACCTGTCGGTGACACCGCT 60
 QY 21 TyrAlaGlnGlnThrArgGlyGluGlyCysGlnGlnThrSerGlnThrGlyArgAsp 40
 DB 61 TACGCTCAGCAGACTCGAGGTGAGCAGGGTTGCCAGAAAGACCTCCACACCGGTGCTGAC 120
 QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
 DB 121 AAAAACACAGGTGAAGGTGAAGTTTCAGATCGTTTCCACCGCTACCCAGACTTCCTGGCT 180
 QY 61 ThrSerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
 DB 181 ACCTCCATCAACGGTGTCTGTCGACCGTTTACACCGTGTGTTACCCGCTACCATCGCT 240
 QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp 100
 DB 241 TCCCGAAAGGTCCGGTTACCCAGATGTACACCAACGTTGACAAAGACCTGGTTGGTTGG 300
 QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
 DB 301 CAGGCTCCGAGGGTTCCCGTTCCCTGACCCCGTGCACCTCGGTTCTCTCGACCTGTAC 360
 QY 121 LeuValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySer 140
 DB 361 CTGTTACCGGTGACGCTGAGCTTATCCCGTTTCGTCGTGCTGCTGCTGCTGCTGCTGCT 420
 QY 141 LeuLeuSerProAcqProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
 DB 421 CTGCTGTCCCGCGTCCGATCTCTACCTGAAAGGTTCTCCGTTGGTCCGCTGCTGCTG 480
 QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180
 DB 481 CCGGCTGTCACGCTGTTGGTATCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
 QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
 DB 541 GCTGTTGACTTCATCCCGTTGAATCTCCCTGGAACACCATGCTGCTGCTGCTGCTGCTG 591

RESULT 3

AAA73331

ID AAA73331 standard: DNA; 594 BP.

XX AC

XX AC

XX DT

XX 19-DEC-2000 (first entry)

XX DE

XX Hepatitis C virus NS4A-NS3 fusion protease coding sequence #4.

XX Hepatitis; NS3 protease; viral replication; chronic liver disease;

KW liver failure; liver cancer; mutant; mutein; ds.

XX OS

OS Hepatitis C virus.

OS Synthetic.

XX FH

FH Key Location/Qualifiers

FT CDS 1..594

FT /*tag= a

FT /product= "NS4A-NS3 fusion protein #4"

XX PN WO200040707-A1.
 XX 13-JUL-2000.
 XX 06-JAN-2000; 2000WO-US00345.
 XX 08-JAN-1999; 99US-0115271.
 XX (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V.
 DR WPI: 2000-465976/40.
 DR P-PSDB; AAB15222.
 PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 PT .
 XX Claim 26; Fig 14; 66pp; English.
 PS The present sequence is the coding sequence for a mutated version of a
 CC fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
 CC protease enzymes. These proteins are both essential for the replication
 CC of the virus, acting to cleave its replicative proteins from the
 CC polyprotein produced from the HCV genome. Inhibitors of the two proteins
 CC should be effective as antiviral treatments of HCV infection. This is
 CC useful as HCV can lead to chronic liver disease such as cirrhosis, liver
 CC failure and liver cancer. The present invention concerns a number of NS3
 CC mutants and NS3-NS4A fusion proteins which can be used to identify
 CC inhibitors of this type, as well as enabling structural studies of the
 CC protease and protease-inhibitor complexes. The protein produced from this
 CC sequence contains the alpha-helix0-1 variant.
 XX SQ Sequence 594 BP; 105 A; 187 C; 155 G; 147 T; 0 other;

Alignment Scores:

Pred. No.: 3.44e-85 Length: 594
 Score: 1002.00 Matches: 194
 Percent Similarity: 98.48% Conservativity: 0
 Best Local Similarity: 98.48% Mismatches: 3
 Query Match: 98.53% Indels: 0
 DB: 21 Gaps: 0

US-09-965-594-18 (1-197) x AAA73331 (1-594)

QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
 DB 1 ATGAAAAAAGAGTCCGTTGTTATCGCGCGCTATCAACCTGTCGGTGACACCGCT 60
 QY 21 TyrAlaGlnGlnThrArgGlyGluGlyCysGlnGlnThrSerGlnThrGlyArgAsp 40
 DB 61 TACGCTCAGCAGACTCGAGGTGAGGAGGGTTGCCAAGAAACCTCCACACCGGTGCTGAC 120
 QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
 DB 121 AAAAACACAGGTGAAGGTGAAGTTTCAGATCGTTTCCACCGCTACCCAGACTTCCTGGCT 180
 QY 61 ThrSerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
 DB 181 ACCTCCATCAACGGTGTCTGTCGACCGTTTACACCGGTGCTGCTGCTGCTGCTGCTGCTG 240
 QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp 100
 DB 241 TCCCGAAAGGTCCGGTTACCCAGATGTACACCAACGTTGACAAAGACCTGGTTGGTTGG 300
 QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
 DB 301 CAGGCTCCGAGGGTTCCCGTTCCCTGACCCCGTGCACCTCGGTTCTCTCGACCTGTAC 360
 QY 121 LeuValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySer 140

```

|||||
361 CTGGTACCCGTCAGCTGACGTTATCCCGTTCGTCGTCGTCACCTCCCGTGGTCC 420
|||||
141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
|||||
421 CTGCTGTCCCGCGCTCCGATCTCTACCTGAAAGGTTCCTCCCGTGGTCCGCTGTGC 480
|||||
161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180
|||||
481 CCGCTGTGTCACGCTGTGGTATCTCCGTCGTCGTCGTCACCGGTGTGTCTAA 540
|||||
181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
|||||
541 GCTGTGACTTCATCCCGGTGAATCCCTGAAACACCATGCTCCCGG 591
|||||
RESULT 4
AAA73334
ID AAA73334 standard; DNA; 594 BP.
AC AAA73334;
XX
XX
DT 19-DEC-2000 (first entry)
XX
DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #7.
XX
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW liver failure; liver cancer; mutant; mutein; ds.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
XX Key Location/Qualifiers
XX CDS 1..594
XX /*tag= a
XX /product= "NS4A-NS3 fusion protein #7"
XX
XX WO200040707-A1.
XX
XX 13-JUL-2000.
XX
XX 06-JAN-2000; 2000WO-US00345.
XX
XX 08-JAN-1999; 99US-0115271.
XX
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX
XX WPI: 2000-465976/40.
XX
XX P-PSDB; AAB15225.
XX
XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
XX substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
XX amino acid, useful for screening inhibitors that may treat hepatitis C
XX
XX
XX Claim 26; Fig 17; 66pp; English.
XX
XX The present sequence is the coding sequence for a mutated version of a
XX fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
XX protease enzymes. These proteins are both essential for the replication
XX of the virus, acting to cleave its replicative proteins from the
XX polyprotein produced from the HCV genome. Inhibitors of the two proteins
XX should be effective as antiviral treatments of HCV infection. This is
XX useful as HCV can lead to chronic liver disease such as cirrhosis, liver
XX failure and liver cancer. The present invention concerns a number of NS3
XX mutants and NS3-NS4A fusion proteins which can be used to identify
XX inhibitors of this type, as well as enabling structural studies of the
XX protease and protease-inhibitor complexes. The protein produced from this
XX sequence contains the alpha-helix0-7 variant.
XX
XX Sequence 594 BP; 105 A; 192 C; 151 G; 146 T; 0 other;

```

```

Alignment Scores:
Pred. No.: 1,56e-84 Length: 594
Score: 995.00 Matches: 193
Percent Similarity: 98.98% Conservative: 2
Best Local Similarity: 97.97% Mismatches: 2
Query Match: 97.84% Indels: 0
DB: 21 Gaps: 0

US-09-965-594-18 (1-197) x AAA73334 (1-594)
QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
DB 1 ATGAAAAAAGGATCCGTTGTTATCGTCGGCGTATCAACCTGTCGGTGACACCGCT 60
QY 21 TyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlnThrSerGlnThrGlyArgAsp 40
DB 61 TACGCTACGACACTCGAGGTGAGGTTACCCGAGNAGACCTCCCNACCCGCTGTGAC 120
QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
DB 121 AAAAACCAAGTTGAAGGTGAAGTTCAGATCGTTTCCACCGCTACCCAGACCTTCCTGGCT 180
QY 61 ThrSerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
DB 181 ACCTCCATCAACGGTGTCTGTGACCGTTTACACCGGTGCTGCTACCCGCTACCATCGCT 240
QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr 100
DB 241 TCCCGAAAGTCGCGTTTACCCAGATGTACACCAACGTTGACAAAGACCTGTTGGTTGG 300
QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
DB 301 CAGGCTCCGACGAGGTTCCTGTCCTGACCCCGTGCACCTCGCGTTCCTCCGACCTGTAC 360
QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
DB 361 CTGGTTACCGCTCAGCTGACGTTATCCCGTTCGTCGTCGTCGTCGTCGTCGTCGTC 420
QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
DB 421 CTGCTGTCCCGCGCTCCGATCTCCTACCTGAAAGGTTCTCCCGTGGTCCGCTGCTGTC 480
QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180
DB 481 CCGGCTGTGTCACGCTGTGGTATCTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 540
QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
DB 541 GCTGTGACTTCATCCCGGTGAATCCCTGGAACACCAACCATGCTGTCCTCCCGG 591

RESULT 5
AAA73330
ID AAA73330 standard; DNA; 594 BP.
XX
XX AAA73330;
XX
XX 19-DEC-2000 (first entry)
XX
XX Hepatitis C virus NS4A-NS3 fusion protease coding sequence #3.
XX
XX Hepatitis; NS3 protease; viral replication; chronic liver disease;
XX liver failure; liver cancer; mutant; mutein; ds.
XX
XX Hepatitis C virus.
XX Synthetic.
XX
XX Key Location/Qualifiers
XX CDS 1..594
XX /*tag= a
XX /product= "NS4A-NS3 fusion protein #3"
XX
XX WO200040707-A1.

```


Percent Similarity: 95.43% Conservative: 1
 Best Local Similarity: 94.92% Mismatches: 7
 Query Match: 93.51% Indels: 2
 DB: 21 Gaps: 1

US-09-965-594-18 (1-197) x AAA73329 (1-588)

QY 1 MetLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
 DB 1 ATGAAAAAAGGATCCGTTTATCGTCGCCGTATAGTACTGAACGGT-----GCT 54
 QY 21 TyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAsp 40
 DB 55 TAGCCTCAGCAGCTCCAGGTGAGGAGGTTCGCAAGAAACCTCCACACCGGTGTCAC 114
 QY 41 LysAsnGlnValGluGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
 DB 115 AAAAACCGAGTTGAAGTGAAGTTCAGATCGTTTCACCGCTGCACACCTTCCTGGCT 174
 QY 61 ThrSerIleAsnGlyValLeuThrThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
 DB 175 ACCTGATCAACGGTGTTCGTCGACCGTTTACACGGTGTGCTGCTACCGTACCATCGCT 234
 QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp 100
 DB 235 TCCCGAAAGGTCGGGTATCCAGATGTACACCAAGCTTGACAAAGACCTGGTTGGTGG 294
 QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
 DB 295 CCGCTCCGACGAGGTTCGCTCCGTCACCGCTGCACCTCGGCTCTCCGACCTGTAC 354
 QY 121 LeuValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySer 140
 DB 355 CTGTTTACCGTCCAGCTGACGTATCCCGGTTCGTCGCTGCTGCTGCTGCTGCTGCTCC 414
 QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
 DB 415 CTGCTGTCCCGGTCGAGTCTCTACCTGAAGGTTCTCCGCTGCTGCTGCTGCTGCTGCT 474
 QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180
 DB 475 CCGCTGTGCACGCTGTGTGTATCTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 534
 QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
 DB 535 GCTGTTGACTTCACTCCCGTTGAATCCCTGGAACCAACCATCGGTTCCTCCCG 585

RESULT 7

AAA73335

ID AAA73335 standard; DNA; 594 BP.

XX AC AAA73335;

XX DT 19-DEC-2000 (first entry)

XX DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #8.

XX KW Hepatitis; NS3 protease; viral replication; chronic liver disease;

XX KW liver failure; liver cancer; mutant; mutein; ds.

XX OS Hepatitis C virus.

XX OS Synthetic.

XX FH Key Location/Qualifiers

FT CDS 1..594

FT /*tag= a

FT /product= "NS4A-NS3 fusion protease protein #8"

XX PN WO200040707-A1.

XX PD 13-JUL-2000.

XX PF 06-JAN-2000; 2000WO-US00345.

XX 08-JAN-1999; 99US-0115271.
 XX (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;

XX WPI; 2000-465976/40.

XX P-PSDB; AAB15226.

XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1

XX substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic

XX amino acid, useful for screening inhibitors that may treat hepatitis C

XX Disclosure; Fig 18; 66pp; English.

XX The present sequence is the coding sequence for a mutated version of a

XX fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A

XX protease enzymes. These proteins are both essential for the replication

XX of the virus, acting to cleave its replicative proteins from the

XX polypeptide produced from the HCV genome. Inhibitors of the two proteins

XX should be effective as antiviral treatments of HCV infection. This is

XX useful as HCV can lead to chronic liver disease such as cirrhosis, liver

XX failure and liver cancer. The present invention concerns a number of NS3

XX mutants and NS3-NS4A fusion proteins which can be used to identify

XX inhibitors of this type, as well as enabling structural studies of the

XX protease and protease-inhibitor complexes. The protein produced from this

XX sequence contains the alpha-helix0 wild-type sequence.

XX Sequence 594 BP; 98 A; 189 C; 153 G; 154 T; 0 other;

Alignment Scores:

Pred. No.: 6.27e-80 Length: 594
 Score: 946.00 Matches: 186
 Percent Similarity: 94.42% Conservative: 0
 Best Local Similarity: 94.42% Mismatches: 11
 Query Match: 93.02% Indels: 0
 DB: 21 Gaps: 0

US-09-965-594-18 (1-197) x AAA73335 (1-594)

QY 1 MetLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20

DB 1 ATGAAAAAAGGATCCGTTTATCGTCGCCGTATAGTACTGAACCTGTCGCGTACACCGCT 60

QY 21 TyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAsp 40

DB 61 TAGCCTCAGCAGCTCGAGTCTGCTGGTGTGCATCATCACCCTCCCTGACCGGTGCTGAC 120

QY 41 LysAsnGlnValGluGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60

DB 121 AAAAACCGAGTTGAAGTGAAGTTCAGATCGTTTCCACCGCTGCTCAGACCTTCCTGGCT 180

QY 61 ThrSerIleAsnGlyValLeuThrThrValTyrHisGlyAlaGlyThrArgThrIleAla 80

DB 181 ACCTGATCAACGGTGTTCGTCGACCGTTTACACGGTGTGCTGCTACCGTACCATCGCT 240

QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp 100

DB 241 TCCCGAAAGGTCGGGTATCCAGATGTACACCAAGCTTGACAAAGACCTGGTTGGTGG 300

QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120

DB 301 CCGGCTCCGACGGTTCCTCCCTGACCGCTGCACCTGGGTTCTCCGACCTGTAC 360

QY 121 LeuValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySer 140

DB 361 CTGTTTACCGTCCAGCTGAGTATCCCGGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420

QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160

DB 421 CTGCTGTCCCGGTCGATCTCTACCTGAAGGTTCTCCCGTGTGCTGCTGCTGCTGCTGCTG 480

QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180
 Db 481 CCGGCTGGTCCACGCTGTGGTATCTCCGCGCGCTGTGGTTCACCGGTGGTGTCTAA 540
 QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
 Db 541 GCTGTTGACTTCATCCCGGTGTAATCCCTGGAAACCAACCATCGGTTCCCGG 591

RESULT 8
 AAA73328
 ID AAA73328 standard: DNA: 588 BP.
 XX
 AC AAA73328;
 XX
 DT 19-DEC-2000 (first entry)
 XX
 DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #1.
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
 KW liver failure; liver cancer; ds.
 XX
 OS Hepatitis C virus.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..588
 FT /tag= a
 FT /product= "NS3-NS4A fusion protein"
 XX
 FN WO200040707-A1.
 XX
 PD 13-JUL-2000.
 XX
 PF 06-JAN-2000; 2000WO-US00345.
 XX
 PR 08-JAN-1999; 99US-0115271.
 XX
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX
 PI Wirttekind M, Weinheimer S, Zhang Y, Goldfarb V;
 XX
 DR WPI: 2000-465976/40.
 DR P-PSDB; AAB15212.
 XX
 PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 PT .
 XX
 PS Disclosure; Fig 10; 66pp; English.
 XX
 CC The present sequence is the coding sequence for a fusion protein created
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
 CC proteins are both essential for the replication of the virus, acting to
 CC cleave its replicative proteins from the polyprotein produced from the
 CC HCV genome. Inhibitors of the two proteins should be effective as
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A
 CC fusion proteins which can be used to identify inhibitors of this type, as
 CC well as enabling structural studies of the protease and
 CC protease-inhibitor complexes.
 XX
 SQ Sequence 588 BP; 97 A; 183 C; 153 G; 155 T; 0 other;

Alignment Scores:
 Pred. No.: 9.58e-77 Length: 588
 Score: 912.00 Matches: 182
 Percent Similarity: 92.89% Conservative: 1
 Best Local Similarity: 92.39% Mismatches: 12
 Query Match: 89.68% Indels: 2
 DB: 21 Gaps: 1

US-09-965-594-18 (1-197) x AAA73328 (1-588)
 QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
 Db 1 ATGAATAAAAGGTTCGGTTGTTATCGTCGGCGGTATAGTACTGAAACGGT-----GCT 54
 QY 21 TyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlnThrSerGlnThrGlyArgAsp 40
 Db 55 TACGCTCAGCAGACTCGAGGTCTCGTGGTTGTCATCATCACTCCCTGACCGGTCTGGTAC 114
 QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
 Db 115 AAAAACCCAGGTTGAAGGTGAAGTTCAGATCGTTCCACCGCTGCTCAGACCTTCCTGGCT 174
 QY 61 ThrSerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
 Db 175 ACCTGCATCAACGGTGTGTTGCTGGACCGTTTACCACCGTGTGTGTACCGGTACCATCGCT 234
 QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp 100
 Db 235 TCCCGAAAGGTCCGGTTATCCAGATGTACACCAACGTTGACAAAGACCTGGTGGTGGTGG 294
 QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
 Db 295 CCGGCTCCGACAGGTTCCTCGTCCACCGCGTGACCTGCGGTTCCTCCGACCTGTAC 354
 QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
 Db 355 CTGTTACCCGCTCAGCTGACGTTATCCCGGTTCGTCGTGCTGCTGCTGCTGCTGCTGCT 414
 QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerGlySerGlyProLeuLeuCys 160
 Db 415 CTGCTGTCCTCCGCTCCGATCTCCTACCTGAAGGTTCTCCGCTGCTGCTGCTGCTGCT 474
 QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180
 Db 475 CCGGCTGCTCAGCTGTTGGTATCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTAAA 534
 QY 181 AlaValAspPheIleProValGluSerLeuGlnThrThrMetArgSerPro 197
 Db 535 GCTGTTGACTTCATCCCGGTGTAATCCCTGGAAACCAACCATCGGTTCCCGG 585

RESULT 9
 ABA95615
 ID ABA95615 standard: DNA: 12734 BP.
 XX
 AC ABA95615;
 XX
 DT 21-MAR-2002 (first entry)
 XX
 DE Chimeric BVDV/HCV NS3-wt sequence.
 XX
 KW Pestivirus; Npro; protease; NS3; screening; ds.
 OS Chimeric - Bovine viral diarrhea virus.
 OS Chimeric - Hepatitis C virus.
 XX
 PN US6326137-B1.
 XX
 PD 04-DEC-2001.
 XX
 PF 25-JUN-1999; 99US-0344456.
 XX
 PR 25-JUN-1999; 99US-0344456.
 XX
 PA (SCHE) SCHERING CORP.
 XX
 PI Hong 2, Lai VCH, Lau JYN;
 XX
 DR WPI: 2002-121103/16.
 XX
 PT Nucleic acid construct encoding chimeric Hepatitis C Virus (HCV)

PT pestivirus genome where the Npro protease gene is replaced with NS3
 PT protease gene, useful for in vivo screening of compounds which inhibit
 PT HCV infection

XX Example 2: Columns 17-28; 20pp; English.

XX The present invention relates to a nucleic acid construct encoding a
 CC chimeric Hepatitis C virus (HCV)-pestivirus genome. The construct
 CC comprises a pestivirus genome where a Npro pestivirus protease gene is
 CC replaced with a gene encoding a functional HCV NS3 protease. Furthermore,
 CC each junction site recognised by the Npro protease is replaced with a
 CC junction site recognised by the HCV NS3 protease. The construct is useful
 CC for screening compounds that inhibit HCV in vivo by inhibiting HCV
 CC protease, where screening may be in cell culture or in an animal model.
 CC The present sequence is a chimeric clone of BVDV (bovine viral diarrhoea
 CC virus)/HCV NS3-wt, which was used to illustrate the present invention.

XX Sequence 12734 BP; 4032 A; 2604 C; 3295 G; 2803 T; 0 other;

Alignment Scores:
 Pred. No.: 3.06e-73 Length: 12734
 Score: 892.50 Matches: 177
 Percent Similarity: 92.82% Conservativity: 4
 Best Local Similarity: 90.77% Mismatches: 11
 Query Match: 87.76% Indels: 3
 DB: 24 Gaps: 1

US-09-965-594-18 (1-197) x ABA95615 (1-12734)

Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
 Db 413 GGTAGTGTGTTTATTTGTTGAGATTGTTTATCTGTTAGTGGTAGTATACGGCGTAC 472
 Qy 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41
 Db 473 GCCCAGCAGAGAGGCGCTCCTAGGCTGAAGATCACCAGTCTGACTGGCGGGACAAA 532
 Qy 42 AsnGlnValGluGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
 Db 533 AACCAAGTGGAGGTGAGGTGAGGATCGTGAACCTGCTACCAACCTTCTCTGCAACG 592
 Qy 62 SerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
 Db 593 TGCATCAATGGGTATGCTGGACTGTCTACACGCGGCGGCAACGACCATCGCATCA 652
 Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
 Db 653 CCCAAGGGTCTCTCATCTCAGATGTATACCAATGTGGACCAAGACCTTGTGGCGTGC 712
 Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
 Db 713 GCTCTCAAGTTCGCGCTCATTCACACCTCGACCTGCGGTCTCGGACCTTTACCTG 772
 Qy 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141
 Db 773 GTTACGAGGCACCGCAGCGCTCATTCGCGTGGCGGCGAGGTGATAGCAGGGGTAGC 832
 Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerClyProLeuLeuCysPro 161
 Db 833 CTTTCGCGCGCGCGCCATTTCCTACCTAAAGGCTCTCTCGGGGGGTCGCTGTGTGCCC 892
 Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
 Db 893 CGGGACACCGCTGGCGCTATTTCAGGGCGCGGTGTGCACCGTGGAGTGCCCAAGCG 952
 Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196
 Db 953 GTGACATTATCCCTGTGGAGACCTTAGACACAACCATGAGATCC 997

RESULT 10

AAx80355

ID AAx80355 standard; cDNA; 1998 BP.

XX

AAx80355;
 07-SEP-1999 (first entry)
 HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:105.

HCV; hepatitis C virus; single chain recombinant complex; linker;
 NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
 hydrophobic domain; covalent complex; detection; inhibitor; ss.
 Hepatitis C virus.
 Synthetic.

WO9928482-A2.

10-JUN-1999.

24-NOV-1998; 98WO-US24528.

28-JUL-1998; 98US-0094331.

28-NOV-1997; 97US-0067315.

(SCHE) SCHERING CORP.

Malcolm BA, Taremi SS, Weber PC, Yao N;

WPI; 1999-385385/32.

New hepatitis C virus covalent complexes

Disclosure; Page 166-169; 21lpp; English.

The present invention describes a covalent hepatitis C virus (HCV)
 NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
 NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
 hydrophobic domain of native HCV NS4A peptide is tethered by the linker
 to the amino terminus of the HCV NS3 protease domain. The present
 sequence encodes an example of the above complex. The covalent
 NS4A-NS3 complexes are useful for structural determination and
 determination of mode of binding of HCV inhibitors by NMR spectroscopy.
 They can also be used for detecting inhibitors of the protease activity,
 the helicase activity and the ATPase activity of NS3. The covalent
 NS4A-NS3 complexes are more soluble, stable and active than the non-
 covalent protease-peptide complexes previously available.

Sequence 1998 BP; 411 A; 595 C; 569 G; 423 T; 0 other;

Alignment Scores:
 Pred. No.: 3.27e-73 Length: 1998
 Score: 881.50 Matches: 167
 Percent Similarity: 93.37% Conservativity: 16
 Best Local Similarity: 85.20% Mismatches: 10
 Query Match: 86.68% Indels: 3
 DB: 20 Gaps: 1

US-09-965-594-18 (1-197) x AAx80355 (1-1998)

Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
 Db 64 GGTCTCTGTTTATTTGTTGTTAGTAATTTTATCTGTTAGTGGTAGTATACGGCGTAC 123
 Qy 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41
 Db 124 TCCCAACAGACGGCGGCGCTTACTTGGTTCAGAGAGACTAGCTTACAGCGCGGACAG 183
 Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
 Db 184 AACCAAGTTCGAGGAGAGGTTCAGGTGGTTTCCACCGCAACACAAATCTCTCTCGCGACC 243
 Qy 62 SerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
 Db 244 TGCCTCAACGGCGTGTGTGGACCGCTTTACCATGGTGGTGTCAAGACCTTAGCCGCG 303

QY	82	ProLysGlyProValThrGlnMetTyThrAsnValAspLysAspLeuValGlyTyrGln	101
Db	304	CAAAAGGGCCCAATCACCAGATGACACTAAATGTGGACCAGCACCTCGTGGCTGGCAG	363
QY	102	AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu	121
Db	364	CGCCCCCGGGGGCGGTTCCTGACACCATGACACCTGTGGAGCTCAGACCTTTACTTG	423
QY	122	ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu	141
Db	424	GTCAGAGACATGCTGACGTCAATCCGTGCGCGCGGGCGGACAGTAGGGGAGCCTG	483
QY	142	LeuSerProArgProIleSerTyLeuLysGlySerSerGlyGlyProLeuLeuCysPro	161
Db	484	CYCTCCCCCAGGCGTGTCTCTACTTGAAGGGCTCTTCGGGTGGTCCACTGCTGCCCT	543
QY	162	AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla	181
Db	544	TCGGGGCAGCGTGGGGCATCTTCGGGCTGCCGTATGCACCCGGGGGTTCGAAGGCG	603
QY	182	ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro	197
Db	604	GTGGACTTTGTGCCGTAGATCCATGGAAACTACTATGGGTCGCG	651
RESULT 11			
ID	AAAX80359 standard; CDNA; 1998 BP.		
AC	AAAX80359;		
XX	07-SEP-1999 (first entry)		
DE	HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:109.		
XX	HCV; hepatitis C virus; single chain recombinant complex; linker;		
KW	NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;		
KW	hydrophobic domain; covalent complex; detection; inhibitor; ss.		
XX	Hepatitis C virus.		
OS	Synthetic.		
XX	WO9928482-A2.		
PN	10-JUN-1999.		
PD	XX		
XX	XX		
PF	24-NOV-1998; 98WO-US24528.		
XX	XX		
PR	28-JUL-1998; 98US-0094331.		
PR	28-NOV-1997; 97US-0067315.		
XX	XX		
PA	{ SCHE } SCHERING CORP.		
XX	XX		
PI	Malcolm BA, Taremi SS, Weber PC, Yao N;		
XX	XX		
DR	WPI; 1999-385385/32.		
XX	XX		
PI	New hepatitis C virus covalent complexes		
XX	XX		
PS	Disclosure; Page 179-182; 21lpp; English.		
XX	XX		
CC	The present invention describes a covalent hepatitis C virus (HCV)		
CC	NS4A-NS3 complex comprising a central hydrophobic domain of native HCV		
CC	NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the		
CC	hydrophobic domain of native HCV NS4A peptide is tethered by the linker		
CC	to the amino terminus of the HCV NS3 protease domain. The present		
CC	sequence encodes an example of the above complex. The covalent		
CC	NS4A-NS3 complexes are useful for structural determination and		
CC	determination of mode of binding of HCV inhibitors by NMR spectroscopy.		
CC	They can also be used for detecting inhibitors of the protease activity,		
CC	the helicase activity and the ATPase activity of NS3. The covalent		
CC	NS4A-NS3 complexes are more soluble, stable and active than the non-		
CC	covalent protease-peptide complexes previously available.		
XX	XX		

```

PF 24-NOV-1998; 98WO-US24528.
XX
PR 28-JUL-1998; 98US-0094331.
PR 28-NOV-1997; 97US-0067315.
XX
XX (SCHE ) SCHERING CORP.
XX
PI MalcolM BA, Taremi SS, Weber PC, Yao N;
XX
DR WPI; 1999-385385/32.
XX
PT New hepatitis C virus covalent complexes
XX
PS Disclosure; Page 160-162; 21lpp; English.
XX
XX The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence encodes an example of the above complex. The covalent
CC NS4A-NS3 complexes are useful for structural determination and
CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.
CC They can also be used for detecting inhibitors of the protease activity,
CC the helicase activity and the ATPase activity of NS3. The covalent
CC NS4A-NS3 complexes are more soluble, stable and active than the non-
CC covalent protease-peptide complexes previously available.
XX
SQ Sequence 1998 BP; 410 A; 596 C; 568 G; 424 T; 0 other;

Alignment Scores:
Pred. No.: 7,77e-73 Length: 1998
Score: 877.50 Matches: 167
Percent Similarity: 92.86% Conservative: 15
Best Local Similarity: 85.20% Mismatches: 11
Query Match: 86.28% Indels: 3
DB: 20 Gaps: 1

DS-09-965-594-18 (1-197) x AAX80353 (1-1998)
QY 5 GlySerValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
DB 64 GTTCTGTTGTTATTGTTGGTAGATTATTTATCTGGTAGTGGTAGTATCAGCGCTAC 123
QY 22 AlaGlnGlnThrArgGlyGluGlyCysGlnGlnThrSerGlnThrGlyArgAspIys 41
DB 124 TCCCAACAGACGCGGGGCGTACTTGGTTCAGATACACTAGCCCTTACAGCGCGGACAA 183
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
DB 184 AACCAGTCCAGGAGAGGTTCCAGTGGTTCCACCGCAACACATCTCTCGCGGACC 243
QY 62 SerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
DB 244 TGGCTCAACGGGGGTGTGTGGACCGCTTACCATGGTGTGGCTCAAGACCTTAGCCGCG 303
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101
DB 304 CCAAGGGCCCAATCACCAGATGTACATAATGTGGACGAGACCTCTCGGTGGCAG 363
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
DB 364 CGCCCCCGGGGGCGGTCCTTGACACCATGCACCTGTGGCAGCTCAGACCTTTACTTG 423
QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
DB 424 GTCACGAGACATCTCAGCTCATTCGGTGCAGCGGGGGGCGACAGTAGGGGAGCGCTG 483
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
DB 484 CTCTCCCCAGGCGCTGTCTCTCTACTTGAAGGGGCTCTTGGGGTGTCCACTGCTGCGCT 543
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181

```

```

DB 544 TCGGGGCACGCTGTGGGCTCTTCGGGCTCCGCTATGCACCCGGGGGTTGGAAGCG 603
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
DB 604 GTGGACTTGTGCGCGTAGAGTCCATGGAAGAACTACTATGCGGTCTCG 651

RESULT 13
AAX80354
ID AAX80354 standard; cDNA; 1998 BP.
XX
XX AAX80354;
XX
DT 07-SEP-1999 (first entry)
XX
DE HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:104.
XX
KW HCV; hepatitis C virus; single chain recombinant complex; linker;
KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
KW hydrophobic domain; covalent complex; detection; inhibitor; ss.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
PN WO928482-A2.
XX
PD 10-JUN-1999.
XX
PF 24-NOV-1998; 98WO-US24528.
XX
PR 28-JUL-1998; 98US-0094331.
PR 28-NOV-1997; 97US-0067315.
XX
XX (SCHE ) SCHERING CORP.
XX
PI MalcolM BA, Taremi SS, Weber PC, Yao N;
XX
DR WPI; 1999-385385/32.
XX
PT New hepatitis C virus covalent complexes
XX
PS Disclosure; Page 163-166; 21lpp; English.
XX
XX The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence encodes an example of the above complex. The covalent
CC NS4A-NS3 complexes are useful for structural determination and
CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.
CC They can also be used for detecting inhibitors of the protease activity,
CC the helicase activity and the ATPase activity of NS3. The covalent
CC NS4A-NS3 complexes are more soluble, stable and active than the non-
CC covalent protease-peptide complexes previously available.
XX
SQ Sequence 1998 BP; 410 A; 596 C; 568 G; 424 T; 0 other;

Alignment Scores:
Pred. No.: 7,77e-73 Length: 1998
Score: 877.50 Matches: 167
Percent Similarity: 92.86% Conservative: 15
Best Local Similarity: 85.20% Mismatches: 11
Query Match: 86.28% Indels: 3
DB: 20 Gaps: 1

US-09-965-594-18 (1-197) x AAX80354 (1-1998)
QY 5 GlySerValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
DB 64 GTTCTGTTGTTATTGTTGGTAGATTATTTATCTGGTAGTGGTAGTATCAGCGCTAC 123
QY 22 AlaGlnGlnThrArgGlyGluGlyCysGlnGlnThrSerGlnThrGlyArgAspIys 41

```



```

Db 124 TCCCAACAGACGGCGGCTTACTTGGTTGCATCAAGACTAGCCTTACAGCGCGGACAG 193
Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
Db 184 AACCAAGTCGAGGAGAGGTTTCAGTGTTCACCGCAACACAAATCTTCTCGGGGACC 243
Qy 62 SerIleAsnGlyValLeuIleThrValTyrHisGlyValAlaGlyThrArgThrIleAlaSer 81
Db 244 TGGGTCAACGGCGTGTTCGACCGTTTACATGGTGTGCTGCTCAAGACCTTAGCGGCG 303
Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101
Db 304 CCAAGGGGCGCAATCACCCAGATGTACACTAAATGTGGACCAAGCACCTCGTGGCTGGCAG 363
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
Db 364 CGCGCCCGCGGGCGGCTTCTTGACACCATGACCTGTGGAGCTCAGACCTTTACTTG 423
Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
Db 424 GTCAGGAGACATGCTGACGTCATTCGGTGGCGGGGGCGGACAGTAGGGGGAGCCTG 483
Qy 142 LeuSerProArgProIleSerTyrLeuGlyGlySerSerGlyGlyProLeuLeuCysPro 161
Db 484 CTCCTCCCGCCAGCGCTGCTCTACTTGAAGGGCTCTTTCGGGTGGTCCACCTGCTCTGCCCT 543
Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
Db 544 TCGGGGACCGCTGTGGGCACTTCCTCGGCTGCGGTATGCACCCGGGGGGTTGCAAGGCG 603
Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
Db 604 GTGACCTTGTGGCCGTAGAGTCCATGGAACACTACTATGCGGTCTCG 651

RESULT 14
ABX15706
ID ABX15706 standard; DNA; 612 BP.
XX
AC ABX15706;
XX
DT 28-MAR-2003 (first entry)
XX
DE Anti-viral synthetic prototoxophore associated DNA sequence.
XX
KW Hepatitis C; ds; viral prototoxophore; anti-viral; tumour;
KW virus; infection; antitumour; toxophore; human immunodeficiency virus;
KW HIV infection; herpes simplex virus; HSV; rhinovirus; NS3 protease.
XX
OS Unidentified.
XX
PN WO200287500-A2.
XX
PD 07-NOV-2002.
XX
PF 26-APR-2002; 2002WO-US13223.
XX
PR 27-APR-2001; 2001US-286893P.
XX
PA (NEWB-) NEWBIOTICS INC.
XX
PI Cathers BE, Neuteboom STC, Shepard HM;
XX
WP1: 2003-167102/16.
XX
Novel synthetic viral prototoxophore for treating viral infections, has
PT toxin moiety incorporated into substrate domain specific for viral
PT enzyme, bound and modified by viral enzyme to get converted into
PT toxophore.
XX
Example 1; Page 62; 66pp; English.
XX
This invention relates to a novel synthetic viral prototoxophore

```

```

CC comprising a toxin moiety operatively incorporated into a substrate
CC domain specific for a viral enzyme. This prototoxophore may be bound
CC and modified by the viral enzyme thus converting it to a toxophore.
CC Also disclosed in the invention is a method for enhancing the anti-viral
CC effect of an antiviral agent, this method comprises contacting a cell,
CC infected with a virus or is susceptible to infection, with a
CC prototoxophore. The invention further comprises an assay to identify
CC anti-viral agents, comprising contacting an infected cell with a
CC candidate agent and comparing the ability of the agent to inhibit the
CC growth or infectivity of the virus in the cell. The prototoxophores
CC of the invention may have virucide or antitumour activity. The
CC prototoxophores of the invention may be useful for reducing or
CC inhibiting viral infectivity, by contacting a cell (e.g. lymphocyte,
CC nerve cell, connective tissue cell, muscle cell or hepatocyte) which is
CC infected with a virus or is susceptible to infection with a virus, with
CC an effective amount of the prototoxophore. The cells are cell lines
CC adapted to long term continuous culture or isolated from a subject.
CC The prototoxophore is also useful for ameliorating the severity of a
CC viral infection in a subject, where the virus is selected from human
CC immunodeficiency virus (HIV), herpes simplex virus (HSV), rhinovirus and
CC hepatitis virus, by administering an effective amount of the
CC prototoxophore to the subject. The prototoxophores of the invention are
CC also useful for treating tumours. The present sequence represents an
CC antiviral prototoxophore associated DNA sequence, this sequence is
CC described as a recombinant NS3/NS4 fusion protein in example 1 of
CC the invention although it is clearly not a protein sequence.
XX
SQ Sequence 612 BP; 120 A; 171 C; 191 G; 130 T; 0 other;

```

Alignment Scores:

```

Pred. No.: 3.39e-73 Length: 612
Score: 874.50 Matches: 175
Percent Similarity: 91.28% Conservative: 3
Best Local Similarity: 89.74% Mismatches: 14
Query Match: 85.99% Indels: 3
DB: 25 Gaps: 1

```

US-09-965-594-18 (1-197) x ABX15706 (1-612)

```

Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
Db 19 GGTATGTCGTTCATTGTGGTAGGATCATTTTGTCCGGTAGTGGTAGTATCAGCGGTAC 78
Qy 22 AlaGlnGlnThrArgGlyGluGlyCysGlnGlnThrSerGlnThrGlyArgAspLys 41
Db 79 GCCAGCAGACAAAGGGCGCTCTAGGGTGCATAATCACACGCTAACTGCGGGGACAAA 138
Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
Db 139 AACCAAGTCGAGGGTAGGTCCAGATTGTGTCACTGTGCCCAACACCTTCTCGCAACG 198
Qy 62 SerIleAsnGlyValLeuIleThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
Db 199 TGCATCAATGGGTGTGCTGGACTGCTACACAGGGCGCGGACGACGACCATCGGTCA 258
Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101
Db 259 CCCAAGGTCCTGTCATCCAGATGTATACCAATGTAGACCAAGACCTTGTGGCTGGCCC 318
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
Db 319 GCTTCGAAGGTACCCGCTCATTCACCTTCGCGCTCCCTCGGACCTTTACCTG 378
Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
Db 379 GTCAGGAGCAGCGCGATGTCATTCCTGCGCGCGGGGTGATAGCAGGGGACGCTG 438
Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
Db 439 CTGTCGCCCGCGCCATTTCTTACTTGAAGGCTCTCGGGGGTCCGCTGTTGTGCCCC 498
Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181

```

Db 499 GCGGGCACCGCTGGCCATATTTAGGCGCGGTGTCCACCCGTGGAGTGGCTAAGGCG 558

Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196
 |||||
 Db 559 GTGAGCTTTATCCCTGTGTGGAAACCTAGAGACACCATGAGGTCC 603

RESULT 15
 AAX80345
 ID AAX80345 standard; cDNA; 651 BP.
 XX
 AC AAX80345;
 XX
 DT 07-SEP-1999 (first entry)
 XX
 DE HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:95.
 XX
 KW HCV; hepatitis C virus; single chain recombinant complex; linker;
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
 KW hydrophobic domain; covalent complex; detection; inhibitor; ss.
 XX
 OS Hepatitis C virus.
 OS Synthetic.
 XX
 PN WO9928482-A2.
 XX
 PD 10-JUN-1999.
 XX
 PF 24-NOV-1998; 98WO-US24528.
 XX
 PR 28-JUL-1998; 98US-0094331.
 PR 28-NOV-1997; 97US-0067315.
 XX
 PA (SCHE) SCHERING CORP.
 XX
 PI Malcolm BA, Taremi SS, Weber PC, Yao N;
 XX
 DR WPI; 1999-385385/32.
 XX
 PT New hepatitis C virus covalent complexes
 XX
 PS Disclosure: Page 147-148; 21lpp; English.
 XX
 CC The present invention describes a covalent hepatitis C virus (HCV)
 CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
 CC to the amino terminus of the HCV NS3 protease domain. The present
 CC sequence encodes an example of the above complex. The covalent
 CC NS4A-NS3 complexes are useful for structural determination and
 CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.
 CC They can also be used for detecting inhibitors of the protease activity,
 CC the helicase activity and the ATPase activity of NS3. The covalent
 CC NS4A-NS3 complexes are more soluble, stable and active than the non-
 CC covalent protease-peptide complexes previously available.
 XX
 SQ Sequence 651 BP; 120 A; 187 C; 200 G; 144 T; 0 other;

Alignment Scores:
 Pred. NO.: 3.66e-73 Length: 651
 Score: 874.50 Matches: 166
 Percent Similarity: 93.33% Conservative: 16
 Best Local Similarity: 85.13% Mismatches: 10
 Query Match: 85.99% Indels: 3
 DB: 20 Gaps: 1

US-09-965-594-18 (1-197) x AAX80345 (1-651)

Qy 5 GlySerValIleValGlyArgGlyAsnLeuSerGlyAsp-----ThrAlaTyr 21
 |||||
 Db 64 GGTTCCTGTGTATGTGTGGTAATATTTATCTGGTAGTGGTAGTACAGCGGCTAC 123
 |||||

Qy 22 AlaGlnGlnThrArgGlyGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41
 ::::|

Db 124 TCCACACACACGCGGGCCCTACTTGGTTCAGAAAGAACTAGCTTACAGCGCGGACAAAG 183

Qy 42 AsnGlnValGlnGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
 |||||
 Db 184 AACCAAGTCTGAGGAGAGGTTCCAGGTGGTTTCCACCGCAACACAAATCTTCTCGCGACC 243
 |||||

Qy 62 SerIleAsnGlyValLeuIleThrValTyrHisGlyValGlyThrArgThrIleAlaSer 81
 ::::|
 Db 244 TCGGTCAACGGCGTGTGTGGACCGTTTACCATGGTGTGGCTCAAGACCTTAGCCGGC 303
 |||||

Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
 |||||
 Db 304 CCAAGGGGCAATACACCCAGATGTACACTAATGTGGACACAGGACCTCGTGGCTGGCAG 363
 |||||

Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
 |||||

Db 364 GCGCCCCCGGGCGCGGTCTTGCACACCTGTGGCAGCTCAGACCTTTACTTG 423
 |||||

Qy 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141
 |||||

Db 424 GTCACGAGACATGCTGACGTCATTCCGGTGCCTGGCGGGGGGACAGTAGGGGAGCCTG 483
 |||||

Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
 |||||

Db 484 CTCTCCCCCAGGCGCTGCTCCTACTTGAAGGGCTCTTCGGGTGGTCCACTGCTCTGCCCT 543
 |||||

Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
 ::::|

Db 544 TCGGGCACGCTGTGGGCTCTTCGGGCTGCCGTATGCACCCGGGGGTTGCGAAGGCG 603
 |||||

Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196
 |||||

Db 604 GTGGACTTTGTGCGCGCTAGAGTCCATGGAACACTACTATCGCGTCT 648

Search completed: August 30, 2003, 19:48:09

Job time : 188.939 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 30, 2003, 19:20:43 ; Search time 1910.31 Seconds
(without alignments)
2506.388 Million cell updates/sec

Title: US-09-965-594-18
Perfect score: 1017
Sequence: 1 MKKKGSVVIGRLNSDFA.....VAKAVDFIPVESLEITMRSP 197

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlip
-O=/sqn2_1/USPTO_spool/US09965594/runat_29082003_151919_28322/app_query.fasta_1.2872
-DB=EST -QFMT=fastcap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09965594 -SCGN_1_12630 &runat_29082003_151919_28322 -NCP0=6 -ICPU=3
-NO_WMAP -LARGESQUEERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop-6
-Fgapext-7 -Ygapop=10 -Ygapext=0.5 -DELOP=6 -DELEXT=7

Database :

EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrn:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	106	10.4	1146	12	BM915803 AGENCOURT
C 2	104.5	10.3	1031	14	CB950999 AGENCOURT
C 3	102.5	10.1	1403	13	BQ926101 AGENCOURT
C 4	100.5	9.9	1199	13	BQ892487 AGENCOURT
C 5	95.5	9.4	502	9	AA036834
C 6	95.5	9.4	701	10	BF863244
C 7	95.5	9.4	984	10	BF304699
C 8	95	9.3	629	10	BG089727
C 9	93.5	9.2	905	13	BD542842
C 10	93.5	9.2	1733	12	BM553374
C 11	93	9.1	528	12	BM402566
C 12	93	9.1	644	29	BX238988
C 13	93	9.1	701	29	B2342381
C 14	93	9.1	1213	13	BU541777
C 15	92.5	9.1	772	29	CC406704
C 16	92.5	9.1	789	29	CC406705
C 17	92.5	9.1	938	13	BQ894657
C 18	91.5	9.0	528	28	AQ620249
C 19	91.5	9.0	1294	13	BQ925457
C 20	91	8.9	580	14	CA728398
C 21	91	8.9	646	12	BG853999
C 22	91	8.9	753	13	BU402910
C 23	91	8.9	865	13	BU219343
C 24	91	8.9	906	13	BX344207
C 25	90.5	8.9	814	11	CNS09179
C 26	90.5	8.9	1141	11	AK080545
C 27	90.5	8.9	1440	12	BM467279
C 28	90	8.8	500	12	BM708007
C 29	90	8.8	569	12	BM825317
C 30	90	8.8	617	10	BE055938
C 31	90	8.8	622	9	AV835401
C 32	90	8.8	631	10	AW961059
C 33	90	8.8	658	12	BM830847
C 34	90	8.8	757	12	B1258851
C 35	90	8.8	763	13	BX093694
C 36	90	8.8	812	13	BO434921
C 37	90	8.8	859	13	BQ222616
C 38	90	8.8	987	13	BX357170
C 39	90	8.8	1015	13	BX404631
C 40	90	8.8	1049	29	CNS040AW
C 41	90	8.8	1123	13	BX398349
C 42	90	8.8	1201	9	AL562877
C 43	90	8.8	1201	13	BX339469
C 44	90	8.8	1564	12	BG827880
C 45	89.5	8.8	409	14	CB805033

ALIGNMENTS

RESULT 1
BM915803/C
LOCUS BM915803
DEFINITION AGENCOURT_6639455 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5482056
5', mRNA sequence.
ACCESSION BM915803
VERSION BM915803.1 GI:19366182
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1146)

QY	147	eSerTyrlEuLySgGlySerSergGlyProLeuLeuCysProAlaGlyHisAlaValGI	167
Dd	786	CGGGTATCTACAGGCCGCAGCCAGCACCATACACTCTCTCCG-----TG	742
QY	167	yIlePheArgAlaAlaValSerThrArgGlyValAlaLysAlaValAspPhe---	186
Dd	741	GGCCTTCGGCGGTCTGGTGTCTTCGCGGTGCGCGGGGGGGGGGGTTTCGCTACC	682
QY	186	vAl 187	
Dd	681	TTTG 678	
RESULT 2			
CB950999			
LOCUS	DEFINITION	CB950999	1031 bp mRNA linear EST 29-APR-2000
ACCESSION	IMAGE:30316162 5', mRNA sequence.		
VERSION	CB950999		
KEYWORDS	EST.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	NIH-MGC http://mgc.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: Dr. Michael Brownstein CDNA Library Preparation: Michael Brownstein Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: NDCM107 row: b column: 11 High quality sequence stop: 333. Location/Qualifiers 1..1031 /organism="Mus musculus" /mol_type="mRNA" /db_xref="taxon:10090" /clone="IMAGE:30316162" /lab_host="DH10B (TI-phage-resistant)" /clone_lib="NIH_MGC_177" /note="Organ: liver; Vector: pDNR-LIB; Site_1: SfiI (ggccatctgcc); Site_2: SfiI (ggccgcctcgcc); CDNA made by oligo-dT priming and directionally cloned. 5' and 3' adaptors were used in cloning as follows: 5'-AAGCAGTGTATCAACGAGTAGTGCATTACGCCGGG-3' and 5'-ATTCTAGACGGGCGGCGGAGATG-DT(30)NN-3'. Full-length enriched library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5 kb size fraction. Library created in the laboratory of M. Brownstein (NIMH, NIH). Note: this is a NIH_MGC Library."		
FEATURES			
source			
BASE COUNT	235 a	309 c	211 g 275 t
ORIGIN			
Alignment Scores:			
Pred. No.:	3.8	Length:	1031
Score:	104.50	Matches:	51
Percent Similarity:	41.10%	Conservative:	16
Best Local Similarity:	31.29%	Mismatches:	62
Query Match:	Indels:	35	
DB:	14	Gaps:	8
US-09-965-594-18 (1-197) x CB950999 (1-1031)			
QY	44	ValGUtUGluValGlInIleValSerThrAlaThrGlnTrPheLeuAlaThrSerIle	63
Dd	395	ATTCAGGATCTCCCACAACAAAGAGAGTACATCGGCAAGCTTTTCCTT-C	ACTCATTT 451

Qy 64 AsnGlyValLeuTTPThrValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLys 83
 Db 452 TTGGGCACACTGTGTCGGTGGACAT-----ATCGATCCCGCTAAA 493

Qy 84 GlyProValThrGlnMetTyrThrAsnValAspLysValGlyTTPGlnAlaPro 103
 Db 494 GGGCCTTTACAAAA-----ACACTTAACCT-CCTTGGCTGGCTGGCATGTGG 543

Qy 104 Gln-----GlySerArgSerLeuThrProCysThrCysGlySerSerAsp 118
 Db 544 CAAAGAACCGTTTTGGGTTCGGCTCTTGGCCCCCCCCCAATTGGGAACCACTGGC 603

Qy 119 LeuTyrLeuValThrArgHisAlaAsp-ValIleProValArgArgGlyAspSer 138
 Db 604 -----ACCACCATGGGCTGTGTTCCTGGCTCTCCGCTGGCAATAC 651

Qy 138 gGlySerLeuSerProArgProfileSerTyrLeuLysGlySerSergly----- 155
 Db 652 AAACNCCCTTAACCGTCCCTCCCAACAATATTCTTCAAGCGTCTCTGATTCCCTAA 711

Qy 156 -GlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaValSerTh 175
 Db 712 GTCCCGCCCTTTTACCCAGACCACTTGTGGGACACAGCGCTCTTTTATCTTC 771

Qy 175 rArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrMetAr 195
 Db 772 C-----CCCTCATGCTCTT---CCACACGGCG 798

Qy 195 gSerPro 197
 Db 799 AACACCC 805

RESULT 3
 BQ926101/c
 LOCUS
 DEFINITION AGNCOURT_8752655 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6335718
 5', mRNA sequence.

ACCESSION BQ926101
 VERSION BQ926101.1 GI:22341132
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1403)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgaaps-r@mail.nih.gov
 Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,
 Ph.D.
 cDNA Library Preparation: ResGen, Invitrogen Corp
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM13798 row: j column: 07
 High quality sequence stop: 101.
 Location/Qualifiers
 1..1403
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:6335718"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_130"
 /note="Organ: otocysts; Vector: pCMV-SPORT6.1.cdbd;
 Site 1: EcorV; Site 2: NotI; Cloned unidirectionally.
 Primer: Oligo dT. Average insert size 1.95 kb.
 Constructed by ResGen, Invitrogen Corp. Note: this is a

BASE COUNT 297 a 521 c 237 g 345 t 3 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 8.97 Length: 1403
 Score: 102.50 Matches: 59
 Percent Similarity: 35.50% Conservative: 12
 Best Local Similarity: 29.50% Mismatches: 67
 Query Match: 10.08% Indels: 62
 DB: 13 Gaps: 11

US-09-965-594-18 (1-197) x BQ926101 (1-1403)

Qy 11 GlyArgIleAsnLeuSerGlyAspThrAlaGlnInThrArgGlyGluGluGly 30
 Db 1378 GGGTGTTCANCGGTTCAGGACAGGTGCGCC---GCACATCGACGCTCGGCCAGACT 1322

Qy 31 CysGlnGluThr-----SerGlnThrGlyArgAspLysAsnGln-----Val 44
 Db 1321 TCTCGGGGGCGGTGGCGCATACCCCGGTGGATCGAGTCCAGCGCGCTTGTATACA 1262

Qy 45 GluGlyGluValGlnIleValSerThrAlaThrGlnInThrPheLeuAlaThrSerIleAsn 64
 Db 1261 GAGGGGAAA-----CAG 1250

Qy 65 GlyValLeuTyrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLys--G 84
 Db 1249 GGGTA---TGTTATCAGCGGCTGGGCGAGTACT-----TCCCTAAAGCG 1205

Qy 84 lyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTTPGlnAlaPro 104
 Db 1204 GCGGCGTGGCAGTATATATACCGAGTGGCAAGCGCACGGCGTGAACGTTGACC 1145

Qy 104 InGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeuValThr 124
 Db 1144 AA---GAGAGCACCTGAGCGCCCTCCCTGTGGGCTGTGATATAACAAATGTGCG 1088

Qy 124 rGHisAlaAspValIleProValArgArgGlyAsp----- 136
 Db 1087 GGCACGGTGTGTGTACTACGCGCAGCGCGCTCCACGCGCTCTCTAACAGCGC 1028

Qy 137 -----SerArgGlySerLeuSerProArgProIle-SerTyrLeuLysGlySerSer 154
 Db 1027 CCGCCCTCCCGCGCAAC-----AGGTAATAATCATATCGCGCGGCGGATTTC 980

Qy 155 Gly-----GlyProLeuLeuCys 160
 Db 979 GCATTCGCGGGAGAGCGCGGTGCGGGGGCGCGCTGCGGCGCTGAGGCGC 920

Qy 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyVal 178
 Db 919 AGGAGAGGC-----GGCGTGTTCGGCGGTGAGGACGAGCGCGGTG 875

RESULT 4
 BQ92487
 LOCUS
 DEFINITION AGNCOURT_8417538 Lupski_sympathetic_trunk Homo sapiens cDNA clone
 IMAGE:6192708 5', mRNA sequence.

ACCESSION BQ92487
 VERSION BQ92487.1 GI:22284501
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 1199)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgaaps-r@mail.nih.gov
 Tissue Procurement: Dr. James R. Lupski

```

871 TCAGGGCGTTTAAAGCCCCCGGCTTCCTCGCGCGGCGGAAGCA 913

Db
RESULT 5
AA036834/c
LOCUS
DEFINITION
IMAGE:471945.5, similar to PIR:A55195 A55195 chordin precursor -
African clawed frog ;, mRNA sequence.
AA036834
AA036834.1 GI:1509872
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 502)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman
, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston
, R., Williamson, A., Wohldmann, P. and Wilson, R.
The WashU-Merck EST Project
Unpublished
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Way, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 300

```

```

FEATURES
  source
    Location/Qualifiers
      1..502
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="GDB:3757947"
        /db_xref="taxon:9606"
        /clone="IMAGE:471945"
        /sex="female"
        /dev_stage="adult"
        /lab_host="DH10B"
        /clone_lib="Soares_pregnant_uterus_MbHPU"
        /note="Organ: uterus; Vector: p7T3-Pac; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was primed with a Not I oligo(dT) primer [5' AACTGGGAAGATTGGCGCGCGCTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p7T3 vector. Library went through one round of normalization. Library constructed by M. Fatima Bonaldo."
BASE COUNT      84 a  141 c  165 g  100 t   12 others
ORIGIN
Alignment Scores:
  Pred. No.:      11.6      Length:      502
  Score:          95.50     Matches:      37
  Percent Similarity: 38.51%   Conservative: 20
  Best Local Similarity: 25.00%  Mismatches:  45
  Query Match:      9.39%     Indels:      46
  DB:                9        Gaps:         6

US-09-965-594-18 (1-197) x AA036834 (1-502)

Qy      14 AsnLeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGlyCysGlnGlu 33
      :::::::::::::::::::::: :::: ::::::::::::::
Db      347 CATCTTGCAGGTATACAGCTCTTCTCCTCAAAANGGGGGGCGACTGAGGGGTGCCAAGCT 288
      :::::::::::::::::::::: :::: ::::::::::::::

Qy      34 ThrSerGlnThrGly-----ArgAspLysAsnGlnValGluGlyClnValGlnIleVal 51
      :::: :::: :::: :::: :::: :::: :::: ::::

Db      287 CTGACTTCTTGGGAACCACTGCCCGACAAAGCGGACCGCGGGGCCCATC----- 237

```

```

Qy 52 SerThrAlaThrGlnThrPheLeuAlaThrSerIleAsnGlyValLeu----- 67
Db 236 -----AGCTGCATGGGTNCNCCAGCTGGGGNG 207
Qy 68 -----TrpThrValThrHisGlyAlaGlyThrArgThrIleAlaSerPro 82
Db 206 GGCCCCCGACCCACCTGTCACAGTGTTCGACGAGTCGGT----- 168
Qy 83 LysGlyProValThrGlnMetThrAsnValAspLysAspLeuValGlyTrpGlnAla 102
Db 167 -----GGGTTGACACGACACNGCTGGCA---CAGGCC 138
Qy 103 ProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuThrVal 122
Db 137 AGCGGGGACACTGCACCTTCTCACAGTCACCTCCAGTCGCCCTTGCAGGTGCAG 78
Qy 123 ThrArgHisAlaAspValIleProValArg-ArgArgGlyAspSerArgGlySerLeuLe 142
Db 77 ACAGCACACTTA-----ATTAGCCAAAGGGGGGCACAAACGGGGTGCCNANCN 30
Qy 142 uSerProArgProIleSerTyr 149
Db 29 GTACCCGTTGCCGCCACGCTTC 8

RESULT 6
BF863244
LOCUS
DEFINITION
963042C02.xl C. reinhardtii CC-1690, Stress condition I, normalized
Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
BF863244
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Chlamydomonas reinhardtii
Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadaeae; Chlamydomonas.

REFERENCE
1 (bases 1 to 701)
Grossman, P., Davies, J., Federspiel, N., Harris, E., Hauser, C.,
Lefebvre, A., McDermott, J.P., Shrager, J., Silflow, C. and Stern, D.
Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants; project phase 3
Unpublished
Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.

FEATURES
source
1..701
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, Stress condition I,
normalized, Lambda Zap II"
/notes="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:
XhoI; This library, constructed by John Davies and Jeffrey
McDermott, combines cDNAs from CC-1690 cells grown to
mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min,
1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr
, 4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was
purified from each sample, pooled and cDNA synthesized.
The cDNA was directionally cloned into Lambda Zap II
(Stratagene) in the EcoRI (5') and XhoI (3') sites.
pBluescript II SK- plasmids were excised from the lambda
ZAP clones by superinfection with ExAssist (Stratagene)
phage. The library was normalized using method 4 described
in Bonaldo et al (1996) Genome Research 6: 791-806."
173 a 213 c 175 g 140 t

BASE COUNT
ORIGIN

```

```

Alignment Scores:
Pred No.: 17.9 Length: 701
Score: 95.50 Matches: 38
Percent Similarity: 38.96% Conservative: 22
Best Local Similarity: 24.68% Mismatches: 63
Query Match: 9.39% Indels: 31
DB: 10 Gaps: 7

US-09-965-594-18 (1-197) x BF863244 (1-701)
Qy 71 TyrHisGlyAlaGlyThrArgThrIleAlaSerProLys-----GlyProVal 86
Db 171 CAGCACCATACCTTGCCTCCTCAGCTGTCTCACACCAAAATTATCCCATACGGGCCACTA 230
Qy 87 ThrGlnMetThrThrAsnValAspLysAspLeuValGlyTrpGlnAlaProGlnGlySer 106
Db 231 ACNAAAGTTACATACACAGG-----AAGGACGAGCGCGCTTGGCCACCCCTTTGGAGCCG 284
Qy 107 ArgSerLeuThrProCysThrCysGlySerSerAspLeuThrValThrArgHisAla 126
Db 285 AGAAGCCGACCGCTGTCTCTGGGTCTATCCGATGCTATGCAATCTCCCGGTATCAG 344
Qy 127 AspValIle-----ProValArgArgArgGlyAspSerArg----- 138
Db 345 GAGATCATTTGCTGCTGCTTTAGTCACCCCAAGAGAGCCTGGGAGTGGSCATTTATAA 404
Qy 139 -----GlySerLeuLeuSerProArgProIleSer---Tyr 149
Db 405 GAAGGGGACGGAATTCGTTTGGGAAAGTGCAGCGCCCAANGTCTGACCAAGTGCTA 464
Qy 150 LeuLysGlySerSerGlyProLeuLeuLeuCysProAlaGlyHisAlaValGlyIlePhe 169
Db 465 CTCGAAAGCAGCAATGGGAGCCTTCGCGGTGTCGGGTGCTCGCTCTTAATGTCTCAG 524
Qy 170 ArgAlaAlaVal-----SerThrArgGlyValAlaLysAla--- 181
Db 525 AAAGAAACCATTTGAGTAGAGAGTGCCTGTTTACCCCGAAGGTGAAGGTCACTCTAT 584
Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArg 195
Db 585 GTGGACCGCATACATCTCGAAGACACACAGGTGCTACGCA 626

RESULT 7
BF304699/c
LOCUS
DEFINITION
601888252Fl NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4122276 5',
mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 984)
NIH-MGC http://mgc.ncbi.nlm.nih.gov/.
AUTHORS
TITLE
NATIONAL INSTITUTES OF HEALTH, MAMMALIAN GENE COLLECTION (MGC)
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCMI005 row: 9 column: 13
High quality sequence stop: 646.
Location/Qualifiers
1..984
/organism="Homo sapiens"
/mol_type="mRNA"

```



```

LOCUS      BU542842                905 bp      mRNA      linear      EST 13-SEP-2002
DEFINITION AGENCOURT_10334715 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6574789
5' mRNA sequence.
ACCESSION  BU542842
VERSION     BU542842.1 GI:22853325
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 905)
            NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS     National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE       Unpublished
JOURNAL
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: DCTD/UTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L1CM2770 row: k column: 13
            High quality sequence stop: 633.
            Location/Qualifiers
                1..905
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:6574789"
                /tissue_type="carcinoma, cell line"
                /lab_host="DH10B (phage-resistant)"
                /clone_lib="NIH_MGC_40"
                /note="organ: prostate; vector: pOTB7; Site_1: XhoI;
                Site_2: EcoRI; cDNA made by oligo-dT priming.
                Directionally cloned into EcoRI/XhoI sites using the
                following 5' adaptor: GGCACGAG(G). Library constructed by
                Ling Hong in the laboratory of Gerald M. Rubin (University
                of California, Berkeley) using ZAP-cDNA synthesis kit
                (Stratagene) and Superscript II RT (Life Technologies).
                Note: this is a NIH_MGC Library."
            BASE COUNT      201 a 260 c 273 g 171 t
            ORIGIN

Alignment Scores:
Pred. No.:      39, 5      Length:      905
Score:          93.50     Matches:     51
Percent Similarity: 31.88%  Conservative: 15
Best Local Similarity: 24.64%  Mismatches: 74
Query Match:     9.19%     Indels:    67
DB:              13       Gaps:      9

US-09-965-594-18 (1-197) x BU542842 (1-905)

Qy      28  GluGluGlyCysGlnGluThrSerGln---ThrGlyArgAspLysAsnGlnValGluGly 46
Db      884 AAGGAAGGGCCCGCAGTCTGTGTTCCAGGAAAGGCGGACCCGAGACCAAGGAGGAGGCG 825
Qy      47  GluValGlnIleValSerThrAlaThrGlnThrPheLeuAla----- 60
Db      824 GGGGCGCTTCTTCAGGCGCTGTGCACAAAGTGTCCCTTGGGGTGCCCGCCCATGGTCCCA 765
Qy      61  -----ThrSerIleAsnGlyValLeuTrpThr----- 69
Db      764 CATTTCGACGATCCGGCAGAACATGTGGTGGTCTTGGCCCGACAGCAGGACAGCC 705
Qy      70  -----ValThrHisGlyAla----- 74
Db      704 AAGTGGGAGGCGGCGATGGTGCACACCTGGGAGGCGCCCTGGTGTGCAGAACAGCAGCCCA 645
Qy      75  -----GlyThrArgThrIle 79
            ::::

```

```

Db      644 CAGTAGCAGCCCATCCAGGAGAACACCACTCGGAGGGCCACAGCCCTCTCGACGCCCTG 585
Qy      80  AlaSerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGly 99
            ::::
Db      584 GCATCCGCCCGCCAGCCCTCCATCTCAGCGGGATGTGCACGGGTGAGACAGGAATGCAGGGA 525
Qy      100 TrpGlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeu 119
            ::::
Db      524 CGTTCGCCCTAGGTCAGCCCTTCATCCGCCCTGTGTGCTGTGTCGATGGTCAAGGTG 465
Qy      120 -----TyrLeuValThrArgHisAlaAsp-----ValIleProValArgArg 133
            ::::
Db      464 CCCTGCCACAGCTGCTGCAAGCCATCCAGGGCTGCTGTGCTCTCTCCAGCTCACACTCT 405
Qy      134 Arg-----GlyAspSerArgGlySerLeuLeuSerPro-----Arg 145
            ::::
Db      404 CGCCTCCAGGGCCAGCCCTTCATCTCTCAGGATCTGGGTAGTCTCTGGGATCTG 345
Qy      146 ProIleSerTyrLeuLysGlySerGlyProLeuLeuCysProAlaGly----- 163
            ::::
Db      344 CCTCAGAAAGGGCTGGCAGGCTTGTCTGCAGGTGCAGCTGTGGCCCTCTCGTCTCTCTG 285
Qy      164 -----HisAlaValGly 167
            ::::
Db      284 CGGGTGGCTACGGTGCAGGG 264

RESULT 10
BM553374/c
LOCUS     BM553374.1 1733 bp      mRNA      linear      EST 20-FEB-2002
DEFINITION AGENCOURT_6558368 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5742981
5' mRNA sequence.
ACCESSION  BM553374
VERSION     BM553374.1 GI:18792049
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1733)
            NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS     National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE       Unpublished
JOURNAL
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L1AM12761 row: p column: 22
            High quality sequence start: 88
            High quality sequence stop: 539.
            Location/Qualifiers
                1..1733
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:5742981"
                /tissue_type="medulla"
                /lab_host="DH10B"
                /clone_lib="NIH_MGC_119"
                /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                Site_2: EcoRV (destroyed); RNA source normal medulla from
                anonymous male age 27. Library is oligo-dT primed and
                directionally cloned (EcoRV site is destroyed upon
                cloning). Average insert size 1.3 kb, insert size range
                0.9-3 kb. Library is normalized and enriched for
                full-length clones and was constructed by C. Gruber
                (Invitrogen). Research Genetics tracking code 013. Note:
                this is a NIH_MGC Library."
            BASE COUNT      243 a 673 c 521 g 288 t 8 others

```

ORIGIN

Alignment Scores:
 Pred. No.: 92.2 Length: 1733
 Score: 93.50 Matches: 56
 Percent Similarity: 37.26% Conservative: 23
 Best Local Similarity: 26.42% Mismatches: 78
 Query Match: 9.13% Indels: 55
 DB: 12 Gaps: 14

US-09-965-594-18 (1-197) x BM553374 (1-1733)

QY 17 GlyAspThrAlaValGlnGlnThrArg-----GlyGluGlu 29
 Db 1027 GGGGACTCAGGGGGGCTCTCTCAGTCGCGCGGGGTCTCCCAAGGGGGGAGAA 968
 QY 30 Gly-----CysGlnGluThrSerGlnThrGlyArgAspLysAsnGlnVal 44
 Db 967 GGGTCTTCAGAGGCTTTGTGAGCGCTTTTGGGAGAGCTAGGCCGAAAGAGAGGTC 908
 QY 45 GluGlyGluValGln-----IleValSerThrAlaThrGlnThrPheLeuAlaThrSer 62
 Db 907 CCGGGGGACCAAAATTCCTGCTGTTGCCAATTCACCAAGGCT---GTTTCATCCAG 851
 QY 63 IleAsnGlyValLeu-----TrpThrValThrHisGlyAlaGlyThrArgThr---Ile 79
 Db 850 GTGGAGGCGCTCTTCGCGGGTGGCCCTTTGGAGGGGAGGCGCGCGGAGTGGTTT 791
 QY 80 AlaSerProLysGlyProValThrGlnMetThrThrAsnValAlaAspLysAspLeuValGly 99
 Db 790 CCTCCCGGAGGAGTCTCCAAAGGGGTTTTTCC-----CTCCAGGC 746
 QY 100 -----TrpAlaProGlnGlySerArgSerLeuThrProCys 112
 Db 745 TCCCGTCCGCGCTTGTCTTGGGAAGCCCGCCAGGCGCAGCA----- 701
 QY 113 ThrCysGlySerSerAspLeuThrValThrArgHisAlaAspValIleProValArg 132
 Db 700 -----GGGCGCGAGTGGGCGAGTGTCTCAGGAGACATGCA----- 668
 QY 133 ArgArgGlyAspSerArgGlySerLeuLeuSerProArg---ProIleSerTyrLeuLys 151
 Db 667 CCGCGGGTGATGGCGCCGCGCGGCTGCGCCATCTTCCCTGGCGCTGGAGAGGCGCAGG 608
 QY 152 GlySerSerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAla 171
 Db 607 GGATGCTCCAGGAGTGCCCGCGGCTGCCAGTGGCTGAGCGTGGAAAGTGCCCGCTGCT 548
 QY 172 AlaValSer-----ThrArgGlyValAlaLys-----AlaValAspPheIle 185
 Db 547 GCTGTGGAACACTGACACATACACAGAGTGGCCCGGTCGCCCGGAGTCTCTGCTCTC 488
 QY 186 ProValGluSerLeuGluThrThrMetArgSerPro 197
 Db 487 CCA-----CGCAGGGGGCGGCGGAGTCCG 464

RESULT 11
 BM402566
 LOCUS 528 bp mRNA linear EST 01-JUL-2002
 DEFINITION SLA005f12_34513 An expressed sequence tag (EST) collection from the resurrection plant *Selaginella lepidophylla* SLA005f12 5, mRNA sequence.

ACCESSION BM402566.1 GI:21643782
 VERSION EST.
 KEYWORDS *Selaginella lepidophylla*
 SOURCE *Selaginella lepidophylla*
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Lycopodiophyta; Isoetopsida; *Selaginellales*; *Selaginellaceae*; *Selaginella*

REFERENCE 1 (bases 1 to 528)
 AUTHORS Iturriaga, G. and Cushman, J.C.
 TITLE An expressed sequence tag (EST) collection from the resurrection

JOURNAL COMMENT
 plant *Selaginella lepidophylla*
 Unpublished
 Contact: Cushman JC
 Department of Biochemistry
 University of Nevada
 MS200, Reno, NV 89557-0014, USA
 Tel: 775-784-1918
 Fax: 775-784-1650
 Email: jcushman@unr.edu
 PCR PRIMERS
 FORWARD: T3 20mer
 BACKWARD: T7 21mer
 Plate: 005 row: F column: 12
 Seq primer: T3 20mer
 High quality sequence stop: 528.

FEATURES
 source
 1..528
 /organism="Selaginella lepidophylla"
 /mol_type="mRNA"
 /db_xref="taxon:59777"
 /clone="SLA005f12"
 /tissue_type="microphyll fronds undergoing desiccation for 2.5 h"
 /dev_stage="adult"
 /clone_lib="An expressed sequence tag (EST) collection from the resurrection plant *Selaginella lepidophylla*"
 /note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site_1: EcoRI; Site_2: XhoI; Library construction was performed according to manufacture's (Stratagene, Inc.) recommended protocol for the Lambda UniZapXR vector and cDNA synthesis kit."
 129 a 125 c 137 g 137 t

BASE COUNT 129 a 125 c 137 g 137 t

ORIGIN

Alignment Scores:
 Pred. No.: 21.9 Length: 528
 Score: 93.00 Matches: 37
 Percent Similarity: 42.98% Conservative: 15
 Best Local Similarity: 30.58% Mismatches: 43
 Query Match: 9.14% Indels: 26
 DB: -12 Gaps: 4

US-09-965-594-18 (1-197) x BM402566 (1-528)

QY 94 AspLysAspLeuValGlyTrpGlnAlaProGlnGlySerArgSerLeuThrProCysThr 113
 Db 53 GACAAGGATGTAGCGGTGCTGAAGATCGATGCTCAAGCAACAGATCTCAGCCCAATACCC 112
 QY 114 CysGlySerSerAspLeuTyrLeuVal----- 122
 Db 113 CTTGGAAGTCTCGCATCTGCTTGTGCCAGAGGTGTATGCTATCGGTAATCCTTTT 172
 QY 123 -----ThrArgHisAlaAspValIleProValArgArgGlyAspSerArg 138
 Db 173 GGATTTGGATCATACGCTGACACAGCGCTCATGCTCTTCCAAGGGGAGATTACT--- 229
 QY 139 GlySerLeuLeuSerProArgProIleSerTyrLeu----- 150
 Db 230 ---TCAGCGCTTATGGTGTCTCCATTCAGACGTGTATCCAGACAGATGCGCGCTATTAT 286
 QY 151 LysGlySerSerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArg 170
 Db 287 CCGTGGACACAGCGGGGTCGCTATTGGACAGTCTTGGAATTTGTAGGATCAACACT 346
 QY 171 AlaAlaValSerThrArgGlyValAlaLysAlaValAspPhe---IleProValGluSer 189
 Db 347 GCTATATATTTCTCCGCTCTGGCGCTTCATCAGCGGTTCATTCATTCAGTTGACACG 406
 QY 190 Leu 190
 Db 407 GTT 409

RESULT 12

```

BX238988/c
LOCUS      BX238988               644 bp    DNA        linear    GSS 29-JAN-2003
DEFINITION Danio rerio genomic clone DREY-283L13, genomic survey sequence.
ACCESSION  BX238988
VERSION    BX238988.1  GI:28161322
KEYWORDS   GSS.
SOURCE     Danio rerio (zebrafish)
ORGANISM   Danio rerio
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE  1 (bases 1 to 644)
AUTHORS    Humphray, S.J., Huckle, E. and Durham, J.L.
TITLE      Direct Submission
JOURNAL    Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humphray@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 283L13. 283L13
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES   Location/Qualifiers
            source             1..644
                                /organism="Danio rerio"
                                /mol_type="genomic DNA"
                                /db_xref="taxon:7955"
                                /clone="DREY-283L13"
                                /tissue_type="Testis"
                                /note="vector pIndigoBAC-536"
BASE COUNT 129 a 212 c 176 g 127 t
ORIGIN
Alignment Scores:
Pred. No.:      28.4      Length:      644
Score:          93.00     Matches:     37
Percent Similarity: 44.72%  Conservative: 18
Best Local Similarity: 30.08%  Mismatches: 52
Query Match:    9.14%      Indels:     16
DB:             29         Gaps:       7

US-09-965-594-18 (1-197) x BX238988 (1-644)
QY 68 TrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLysGlyProValThr 87
DB 405 TGGACAGCATCTCTGGCAGAGTGGAGGCTCGCAGGGCGGTGTCGCCGTCCACGATGACC 346
QY 88 GlnMetTyr-----ThrAsnValAspLysAspLeuValGlyTrpGlnAla 102
DB 345 AGCAGCTGGTCATCGTGGTGTCCAGCAGTCTTGGAAAGGAC-----TGGAGATCC 295
QY 103 ---ProGlnGlySerArg-----SerLeuThrProCysThrCysGlySerSerAsp 118
DB 294 AGACCGTTGGGGACAAGAGCGGAGGAGTGTGTCACCATGCGCTTTAAAAAAGGAGAAA 235
QY 119 LeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArg 138
DB 234 TTATTGTGATTACCATGGGAGGAGTGTGCGCAGGAGGCGCTTAGGAGAGCGGAGG 175
QY 139 GlySerLeuLeuSerProArgPro-----IleSerTyrLeuLysGlySerSerGlyGly 156
DB 174 CCTCTCCCGGTGTCCTCTCTCTCATATGTTATTTCT---TTAAAGGCTTGGTGGGAGGA 118
QY 157 ProLeuLeu---CysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThr 175
DB 117 CCTTTGCTGATGCCCGACGCTTGGCCCTCTGGCCCTCTGTCACCCGGGCATGGAACCTTCG 58
QY 176 ArgGlyVal 178
DB 57 GAAGCGCTTA 49

RESULT 13
BX242381/c
LOCUS      BX242381               701 bp    DNA        linear    GSS 06-NOV-2002

```

```

DEFINITION ic83b11.b1 WGS-SbicolorF (JM107 adapted methyl filtered) Sorghum
            bicolor genomic clone ic83b11 5', genomic survey sequence.
ACCESSION  BZ342381
VERSION    BZ342381.1  GI:24742983
KEYWORDS   GSS.
SOURCE     Sorghum bicolor (sorghum)
ORGANISM   Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE  1 (bases 1 to 701)
AUTHORS    Rabinowicz, P.D., O'Shaughnessy, A.L., Balija, V., Dedhia, N.,
            Katzenburger, F., King, L., Miller, B., Muller, S., Nascimento, L.,
            Zutavern, T., Palmer, L., McCombie, W.R. and Martienssen, R.A.
            Genomic shotgun sequences from Sorghum bicolor (methyl-filtered)
            Unpublished
            Contact: W. Richard McCombie
            Lita Annenberg Hazen Genome Sequencing Center
            Cold Spring Harbor Laboratory
            PO Box 100, Cold Spring Harbor, NY 11724, USA
            Tel: 516 367 8884
            Fax: 516 367 8874
            Email: mcombie@cshl.org
            Plate: ic83 row: b column: 11
            Seq primer: -21M13UnivFwd
            Class: shotgun
            High quality sequence stop: 701.
FEATURES   Location/Qualifiers
            source             1..701
                                /organism="Sorghum bicolor"
                                /mol_type="genomic DNA"
                                /db_xref="taxon:4558"
                                /clone="ic83b11"
                                /lab_host="JM107 or DH5a"
                                /clone_lib="WGS-SbicolorF (JM107 adapted methyl filtered)"
                                /note="Site 1: Xba I; Site 2: Xba I; The vector was
                                digested with XbaI and one nucleotide was added by fill in
                                in the recessive 3' end. The genomic DNA was nebulized,
                                end repaired, adaptor ligated and size fractionated using
                                sephadex. The resulting fragments were between 0.8 and 3
                                kb and were cloned into the vector (-x/y reads in M13mp19,
                                .b/g reads in pUC19). The same ligation was transformed in
                                either JM107 or DH5a."
BASE COUNT 108 a 251 c 232 g 110 t
ORIGIN
Alignment Scores:
Pred. No.:      31.7      Length:      701
Score:          93.00     Matches:     48
Percent Similarity: 36.69%  Conservative: 14
Best Local Similarity: 28.40%  Mismatches: 56
Query Match:    9.14%      Indels:     51
DB:             29         Gaps:       8

US-09-965-594-18 (1-197) x BZ342381 (1-701)
QY 60 AlaThrSerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThr--- 78
DB 500 GCCCATGCCGTGGCGGTGTTTTTTTGACGTACCGCACGGCGGGGAGGACGACGAGC 441
QY 79 -----IleAlaSerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAsp 96
DB 440 GTGGCGGTGGCCCGCCCGCCCGCGCGGTGGGAATATGCGCCCAATAAATCTCGCAC 381
QY 97 LeuValGlyTrpGlnAlaProGlnGlySerArg----SerLeuThrProCysThr----- 113
DB 380 CTACGCGGAGCGGAGAGAGAGAGGTGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 321
QY 114 -----CysGlySerSerAsp 118
DB 320 CACCCAGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 261
QY 119 LeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArgGlyAsp----- 136

```


ORIGIN

Alignment Scores: 40.3 Length: 772
Pred. No.: 92.50 Matches: 47
Score: 40.57% Conservative: 24
Percent Similarity: 40.57% Mismatches: 59
Best Local Similarity: 26.86% Indels: 46
Query Match: 9.10% Gaps: 10
DB: 29

US-09-965-594-18 (1-197) x CC406704 (1-772)

```
QY 39 ArgAspLysAsnGlnVal---GluGlyGluValGlnIleValSerThrAlaThrGlnThr 57
Db 633 AGATGGCGAAGCAGCTAAACAGGGCTATGTACAAATTGTGACG-----AGCCTATG 580
QY 58 PheLeuAlaThrSerIleAsnGlyVal---LeuTrpThrValTyrHisGlyAlaGlyThr 76
Db 579 TACGAGGCCACCGCTGTCCCGCTATTCTATCTGGAGG-----TCA 538
QY 77 ArgThrIleAlaSerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAsp 96
Db 537 CAGACTTTCTGACGATGATCAAGGTGACACAGATGATGACCAGG-----AAGAGC 484
QY 97 LeuValGlyTyrP---GlnAlaProGlnGlySerArgSerLeuThrProCysThr 113
Db 483 AGGACCCAGTGTTCCTTACATGCAATAACTGGGATTAGAGGGAGGACACCATGCGAGC 424
QY 114 CysGlySerSerAspLeu-----TyrLeuValThrArgHisAlaAsp 127
Db 423 TCGGGGTAGTCTCTCAACGGTCAAGGCTGCTGGCCCTACTTGACACGGGTTCAACACATA 364
QY 128 ValIleProValArgArgGlyAspSerArgGlySerLeuLeuSerProArgProIle 147
Db 363 ACTTCATCAACTGCACAGCGCGCA-CAACAGCGTGGGGTTACTTTGGAAACCCACACAGGT 305
QY 148 SerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGly 167
Db 304 CGCCATGTCAAGGTGGCAAAATGGAGACCCAGTTTCTGCCAG-----GGA 260
QY 168 IlePheArgAlaAlaValSerThrArgGlyValAlaLysAlaValAspPhe----- 184
Db 259 GTAACTCGTCGGCA-----GCCATTGACATCAACAAGGAG 224
QY 185 -----IleProValGluSerLeuGlu 191
Db 223 AAGTTCACCATTCAGGCATATGCAATTCCTTGGATACATTGAG 179
```

Search completed: August 31, 2003, 04:27:41
Job time : 1917.31 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 17:42:58 ; Search time 44.6227 Seconds
(without alignments)
700.745 Million cell updates/sec

Title: US-09-965-594-20

Perfect score: 1020

Sequence: 1 HKKKGSVIVGRINLSGDTA.....VAKAVDFIPVESLETTMRSP 197

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 15872573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

```

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1020	100.0	197	AA15224	Hepatitis C virus
2	1010	99.0	197	AA15225	Hepatitis C virus
3	1005	98.5	197	AA15223	Hepatitis C virus
4	990	97.1	197	AA15222	Hepatitis C virus
5	973	95.4	197	AA15221	Hepatitis C virus
6	946	92.7	197	AA15226	Hepatitis C virus
7	939	92.1	195	AA15220	Hepatitis C virus
8	912	89.4	195	AA15212	Hepatitis C virus
9	885.5	86.8	665	AA1524943	HCV NS4A-NS3 compl

10	882.5	86.5	665	20	AA1524947	HCV NS4A-NS3 compl
11	881.5	86.4	665	20	AA1524942	HCV NS4A-NS3 compl
12	878.5	86.1	216	20	AA1517880	HCV NS4A-NS3 compl
13	878.5	86.1	665	20	AA1524946	HCV NS4A-NS3 compl
14	877.5	86.0	665	20	AA1524941	HCV NS4A-NS3 compl
15	875.5	85.8	216	20	AA1517884	HCV NS4A-NS3 compl
16	874.5	85.7	216	20	AA1517879	HCV NS4A-NS3 compl
17	874.5	85.7	665	20	AA1524945	HCV NS4A-NS3 compl
18	873.5	85.6	665	20	AA1524940	HCV NS4A-NS3 compl
19	873.5	85.6	671	20	AA1524948	HCV NS4A-NS3 compl
20	871.5	85.4	216	20	AA1517883	HCV NS4A-NS3 compl
21	870.5	85.3	216	20	AA1517878	HCV NS4A-NS3 compl
22	870.5	85.3	665	20	AA1524944	HCV NS4A-NS3 compl
23	870.5	85.3	671	20	AA1524949	HCV NS4A-NS3 compl
24	870	85.3	215	20	AA1517890	HCV NS4A-NS3 compl
25	867.5	85.0	216	20	AA1517882	HCV NS4A-NS3 compl
26	867.5	85.0	216	20	AA1517885	HCV NS4A-NS3 compl
27	866.5	85.0	216	20	AA1517877	HCV NS4A-NS3 compl
28	864	84.7	215	20	AA1517881	HCV NS4A-NS3 compl
29	863.5	84.7	215	20	AA1517885	HCV NS4A-NS3 compl
30	863.5	84.7	216	20	AA1517888	HCV NS4A-NS3 compl
31	859	84.2	213	20	AA1517888	HCV NS4A-NS3 compl
32	859	84.2	631	20	AA153482	HCV NS3 protein.
33	858.5	84.2	191	21	AA1544728	Hepatitis C virus
34	858.5	84.2	3011	19	AA1577397	Hepatitis C virus
35	858.5	84.2	3011	24	ABP71460	Amino acid sequenc
36	858.5	84.2	3012	23	AAU99289	Hepatitis C virus
37	855.5	83.9	3011	14	AA140120	HCV genomic amino
38	854.5	83.8	687	16	AA1579223	phcV150-encoded se
39	854.5	83.8	1648	16	AA1579221	phcV176-encoded se
40	854.5	83.8	1766	10	AA152041	Sequence encoded i
41	854.5	83.8	1786	10	AA1520158	Protein sequence o
42	854.5	83.8	2261	10	AA1520164	Peptide encoded by
43	854.5	83.8	2301	10	AA1520167	Sequence encoded i
44	854.5	83.8	2436	10	AA152050	Sequence encoded i
45	854.5	83.8	2436	10	AA1520288	Peptide encoded by

ALIGNMENTS

```

RESULT 1
AA15224
ID  AA15224 standard; protein; 197 AA.
XX
AC  AA15224;
XX
DT  19-DEC-2000 (first entry)
XX
DE  Hepatitis C virus NS4A-NS3 fusion protease #6.
XX
Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW  liver failure; liver cancer; mutant; mutain.
XX
OS  Hepatitis C virus.
OS  Synthetic.
PN  WO200040707-A1.
XX
PD  13-JUL-2000.
XX
PF  06-JAN-2000; 2000WO-US00345.
XX
PR  08-JAN-1999; 99US-0115271.
XX
PA  (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
PI  Wittekand M, Weinheimer S, Zhang Y, Goldfarb V;
XX
WPI; 2000-465976/40.
DR  N-PSDB; AAA73333.
XX
PT  Modified hepatitis C virus (HCV) NS3 protease comprising at least 1

```

PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 PT
 XX
 PS Claim 23; Fig 16; 66pp; English.
 XX
 CC The present sequence is a mutated version of a fusion protein created
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
 CC proteins are both essential for the replication of the virus, acting to
 CC cleave its replicative proteins from the polyprotein produced from the
 CC HCV genome. Inhibitors of the two proteins should be effective as
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A
 CC fusion proteins which can be used to identify inhibitors of this type, as
 CC well as enabling structural studies of the protease and
 CC protease:inhibitor complexes. This sequence contains the alpha-helix0-7
 CC variant.
 XX
 SQ Sequence 197 AA;

Query Match 100.0%; Score 1020; DB 21; Length 197;
 Best Local Similarity 100.0%; Pred. No. 3.2e-98;
 Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKKGSVVIVGRINLSGDTAYAQOTRGEGCGCKTSHTGRKNOVEGEVQIVSTATOTFLA 60
 DB 1 MKKGSVVIVGRINLSGDTAYAQOTRGEGCGCKTSHTGRKNOVEGEVQIVSTATOTFLA 60
 QY 61 TSINGVLWTVYHGAGTRTIAAPKGPVTOMYTNVDKDLVGVQAPQGSRLTPCTCGSSDLY 120
 DB 61 TSINGVLWTVYHGAGTRTIAAPKGPVTOMYTNVDKDLVGVQAPQGSRLTPCTCGSSDLY 120
 QY 121 LVTRHADVIPVRRGDSRGSLLSPRPISYLGKSSGGPLLCPAGHAVGIFRAAVSTRGVAK 180
 DB 121 LVTRHADVIPVRRGDSRGSLLSPRPISYLGKSSGGPLLCPAGHAVGIFRAAVSTRGVAK 180
 QY 181 AVDFIPVESLETTMRSP 197
 DB 181 AVDFIPVESLETTMRSP 197

RESULT 2
 AAB15225
 ID AAB15225 standard; protein; 197 AA.
 AC AAB15225;
 DT 19-DEC-2000 (first entry)
 XX Hepatitis C virus NS4A-NS3 fusion protease #7.
 DE Hepatitis; NS3 protease; viral replication; chronic liver disease;
 KW liver failure; liver cancer; mutant; muten.
 XX Hepatitis C virus.
 OS Synthetic.
 XX WO200040707-A1.
 XX 13-JUL-2000.
 XX 06-JAN-2000; 2000WO-US00345.
 XX 08-JAN-1999; 99US-0115271.
 XX (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
 PI WPI: 2000-465976/40.
 DR N-PSDB; AAA73334.
 DR

PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 PT
 XX
 PS Claim 23; Fig 17; 66pp; English.
 XX
 CC The present sequence is a mutated version of a fusion protein created
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
 CC proteins are both essential for the replication of the virus, acting to
 CC cleave its replicative proteins from the polyprotein produced from the
 CC HCV genome. Inhibitors of the two proteins should be effective as
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A
 CC fusion proteins which can be used to identify inhibitors of this type, as
 CC well as enabling structural studies of the protease and
 CC protease:inhibitor complexes. This sequence contains the alpha-helix0-7
 CC variant.
 XX
 SQ Sequence 197 AA;

Query Match 99.0%; Score 1010; DB 21; Length 197;
 Best Local Similarity 99.5%; Pred. No. 3.6e-97;
 Matches 196; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MKKGSVVIVGRINLSGDTAYAQOTRGEGCGCKTSHTGRKNOVEGEVQIVSTATOTFLA 60
 DB 1 MKKGSVVIVGRINLSGDTAYAQOTRGEGCGCKTSHTGRKNOVEGEVQIVSTATOTFLA 60
 QY 61 TSINGVLWTVYHGAGTRTIAAPKGPVTOMYTNVDKDLVGVQAPQGSRLTPCTCGSSDLY 120
 DB 61 TSINGVLWTVYHGAGTRTIAAPKGPVTOMYTNVDKDLVGVQAPQGSRLTPCTCGSSDLY 120
 QY 121 LVTRHADVIPVRRGDSRGSLLSPRPISYLGKSSGGPLLCPAGHAVGIFRAAVSTRGVAK 180
 DB 121 LVTRHADVIPVRRGDSRGSLLSPRPISYLGKSSGGPLLCPAGHAVGIFRAAVSTRGVAK 180
 QY 181 AVDFIPVESLETTMRSP 197
 DB 181 AVDFIPVESLETTMRSP 197

RESULT 3
 AAB15223
 ID AAB15223 standard; protein; 197 AA.
 AC AAB15223;
 DT 19-DEC-2000 (first entry)
 XX Hepatitis C virus NS4A-NS3 fusion protease #5.
 DE Hepatitis; NS3 protease; viral replication; chronic liver disease;
 KW liver failure; liver cancer; mutant; muten.
 XX Hepatitis C virus.
 OS Synthetic.
 XX WO200040707-A1.
 XX 13-JUL-2000.
 XX 06-JAN-2000; 2000WO-US00345.
 XX 08-JAN-1999; 99US-0115271.
 XX (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
 PI WPI: 2000-465976/40.
 DR N-PSDB; AAA73332.
 DR

XX
PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
PT amino acid, useful for screening inhibitors that may treat hepatitis C
XX
PS
XX Claim 23; Fig 15; 66pp; English.
XX
CC The present sequence is a mutated version of a fusion protein created
CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
CC proteins are both essential for the replication of the virus, acting to
CC cleave its replicative proteins from the polyprotein produced from the
CC HCV genome. Inhibitors of the two proteins should be effective as
CC antiviral treatments of HCV infection. This is useful as HCV can lead to
CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
CC The present invention concerns a number of NS3 mutants and NS3-NS4A
CC fusion proteins which can be used to identify inhibitors of this type, as
CC well as enabling structural studies of the protease and
CC protease:inhibitor complexes. This sequence contains the alpha-helix0-1
XX variant.
XX
SQ Sequence 197 AA;

Query Match 98.58; Score 1005; DB 21; Length 197;
Best Local Similarity 98.58; Pred. No. 1.2e-96;
Matches 194; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 1 MKKGSVVIVGRINLSGDTAYAAQTGEGCGCKTSHTGRDKNOVEGEVQIVSTATQTFLA 60
1 MKKGSVVIVGRINLSGDTAYAAQTGEGCGCKTSHTGRDKNOVEGEVQIVSTATQTFLA 60
Qy 61 TSINGVLTWVYHGAGTRTITASPKGPVTQMTNVNVDKLVGWAQPGSRSLTPCTCGSSDLY 120
Db 61 TSINGVLTWVYHGAGTRTITASPKGPVTQMTNVNVDKLVGWAQPGSRSLTPCTCGSSDLY 120
Qy 121 LVTRHADVIPVRRGDSRGLSPRPISYLKSGSGGPLCPAGHAGVIFRAAVSTRGVAK 180
Db 121 LVTRHADVIPVRRGDSRGLSPRPISYLKSGSGGPLCPAGHAGVIFRAAVSTRGVAK 180
Qy 181 AVDFIPVESLETTMRSP 197
Db 181 AVDFIPVESLETTMRSP 197

RESULT 4
AAB15222
ID AAB15222 standard; protein; 197 AA.
XX
AC AAB15222;
XX
DT 19-DEC-2000 (first entry)
XX
DE Hepatitis C virus NS4A-NS3 fusion protease #4.
XX
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW liver failure; liver cancer; mutant; mutein.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
PN WO2000040707-A1.
XX
PD 13-JUL-2000.
XX
PF 06-JAN-2000; 2000WO-US00345.
XX
PR 08-JAN-1999; 99US-0115271.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX WPI; 2000-465976/40.

DR N-PSDB; AAA73331.
XX
PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
PT amino acid, useful for screening inhibitors that may treat hepatitis C
XX
PS
XX Claim 23; Fig 14; 66pp; English.
XX
CC The present sequence is a mutated version of a fusion protein created
CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
CC proteins are both essential for the replication of the virus, acting to
CC cleave its replicative proteins from the polyprotein produced from the
CC HCV genome. Inhibitors of the two proteins should be effective as
CC antiviral treatments of HCV infection. This is useful as HCV can lead to
CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
CC The present invention concerns a number of NS3 mutants and NS3-NS4A
CC fusion proteins which can be used to identify inhibitors of this type, as
CC well as enabling structural studies of the protease and
CC protease:inhibitor complexes. This sequence contains the alpha-helix0-1
XX variant.
XX
SQ Sequence 197 AA;

Query Match 97.18; Score 980; DB 21; Length 197;
Best Local Similarity 97.08; Pred. No. 4.4e-95;
Matches 191; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
Qy 1 MKKGSVVIVGRINLSGDTAYAAQTGEGCGCKTSHTGRDKNOVEGEVQIVSTATQTFLA 60
Db 1 MKKGSVVIVGRINLSGDTAYAAQTGEGCGCKTSHTGRDKNOVEGEVQIVSTATQTFLA 60
Qy 61 TSINGVLTWVYHGAGTRTITASPKGPVTQMTNVNVDKLVGWAQPGSRSLTPCTCGSSDLY 120
Db 61 TSINGVLTWVYHGAGTRTITASPKGPVTQMTNVNVDKLVGWAQPGSRSLTPCTCGSSDLY 120
Qy 121 LVTRHADVIPVRRGDSRGLSPRPISYLKSGSGGPLCPAGHAGVIFRAAVSTRGVAK 180
Db 121 LVTRHADVIPVRRGDSRGLSPRPISYLKSGSGGPLCPAGHAGVIFRAAVSTRGVAK 180
Qy 181 AVDFIPVESLETTMRSP 197
Db 181 AVDFIPVESLETTMRSP 197

RESULT 5
AAB15221
ID AAB15221 standard; protein; 197 AA.
XX
AC AAB15221;
XX
DT 19-DEC-2000 (first entry)
XX
DE Hepatitis C virus NS4A-NS3 fusion protease #3.
XX
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW liver failure; liver cancer; mutant; mutein.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
PN WO2000040707-A1.
XX
PD 13-JUL-2000.
XX
PF 06-JAN-2000; 2000WO-US00345.
XX
PR 08-JAN-1999; 99US-0115271.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX

DR WPI: 2000-465976/40.
 DR N-PSDB; AAA73330.
 XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 PT -
 XX Claim 23; Fig 13; 66pp; English.
 XX The present sequence is a mutated version of a fusion protein created
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
 CC proteins are both essential for the replication of the virus, acting to
 CC cleave its replicative proteins from the polyprotein produced from the
 CC HCV genome. Inhibitors of the two proteins should be effective as
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A
 CC fusion proteins which can be used to identify inhibitors of this type, as
 CC well as enabling structural studies of the protease and
 CC protease:inhibitor complexes. This sequence contains the alpha-helix0-1
 CC variant.
 XX SQ Sequence 197 AA:
 Query Match 95.4%; Score 973; DB 21; Length 197;
 Best Local Similarity 95.4%; Pred. No. 2.6e-93; Indels 0; Gaps 0;
 Matches 188; Conservative 2; Mismatches 7;
 QY 1 MKKGSVVIVGRINLSGDTAYAAQTGEGCQKTSHTGRDKNOVEGEVQIVSTATQTFLA 60
 DB 1 MKKGSVVIVGRINLSGDTAYAAQTGEGCQKTSHTGRDKNOVEGEVQIVSTATQTFLA 60
 QY 61 TSINGVLVTYVHGAGTRTIASPKGPVQMTNVDKDLVGHQAPQGSRSLLPTCTCGSSDLY 120
 DB 61 TCINGVCVTVYHGAGTRTIASPKGPVQMTNVDKDLVGHQAPQGSRSLLPTCTCGSSDLY 120
 QY 121 LVTRHADVIPVRRGDSRGSLLSPRPISYLGSSGGPLLCPAGHAVGIFRAAVSTRGVAK 180
 DB 121 LVTRHADVIPVRRGDSRGSLLSPRPISYLGSSGGPLLCPAGHAVGIFRAAVSTRGVAK 180
 QY 181 AVDFIPVESLETTMRSP 197
 DB 181 AVDFIPVESLETTMRSP 197
 RESULT 6
 AAB15226
 ID AAB15226 standard; protein; 197 AA.
 AC AAB15226;
 DT 19-DEC-2000 (first entry)
 DE Hepatitis C virus NS4A-NS3 fusion protease #8.
 XX Hepatitis; NS3 protease; viral replication; chronic liver disease;
 KW liver failure; liver cancer; mutant; mutein.
 XX Hepatitis C virus.
 OS Synthetic.
 XX WO200040707-A1.
 PN 13-JUL-2000.
 XX 06-JAN-2000; 2000WO-US00345.
 XX 08-JAN-1999; 99US-0115271.
 XX (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;

XX WPI: 2000-465976/40.
 DR N-PSDB; AAA73335.
 XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 PT -
 XX Example 5; Fig 18; 66pp; English.
 XX The present sequence is a mutated version of a fusion protein created
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
 CC proteins are both essential for the replication of the virus, acting to
 CC cleave its replicative proteins from the polyprotein produced from the
 CC HCV genome. Inhibitors of the two proteins should be effective as
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A
 CC fusion proteins which can be used to identify inhibitors of this type, as
 CC well as enabling structural studies of the protease and
 CC protease:inhibitor complexes. This sequence contains the alpha-helix0
 CC wild-type sequence.
 XX SQ Sequence 197 AA:
 Query Match 92.7%; Score 946; DB 21; Length 197;
 Best Local Similarity 94.4%; Pred. No. 1.7e-90;
 Matches 186; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
 QY 1 MKKGSVVIVGRINLSGDTAYAAQTGEGCQKTSHTGRDKNOVEGEVQIVSTATQTFLA 60
 DB 1 MKKGSVVIVGRINLSGDTAYAAQTGEGCQKTSHTGRDKNOVEGEVQIVSTATQTFLA 60
 QY 61 TSINGVLVTYVHGAGTRTIASPKGPVQMTNVDKDLVGHQAPQGSRSLLPTCTCGSSDLY 120
 DB 61 TCINGVCVTVYHGAGTRTIASPKGPVQMTNVDKDLVGHQAPQGSRSLLPTCTCGSSDLY 120
 QY 121 LVTRHADVIPVRRGDSRGSLLSPRPISYLGSSGGPLLCPAGHAVGIFRAAVSTRGVAK 180
 DB 121 LVTRHADVIPVRRGDSRGSLLSPRPISYLGSSGGPLLCPAGHAVGIFRAAVSTRGVAK 180
 QY 181 AVDFIPVESLETTMRSP 197
 DB 181 AVDFIPVESLETTMRSP 197
 RESULT 7
 AAB15220
 ID AAB15220 standard; protein; 195 AA.
 AC AAB15220;
 DT 19-DEC-2000 (first entry)
 DE Hepatitis C virus NS4A-NS3 fusion protease #2.
 XX Hepatitis; NS3 protease; viral replication; chronic liver disease;
 KW liver failure; liver cancer; mutant; mutein.
 XX Hepatitis C virus.
 OS Synthetic.
 XX WO200040707-A1.
 PN 13-JUL-2000.
 XX 06-JAN-2000; 2000WO-US00345.
 XX 08-JAN-1999; 99US-0115271.
 XX (BRIM) BRISTOL-MYERS SQUIBB CO.

PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
 XX
 DR WPI: 2000-465976/40.
 DR N-PSDB; AAA73329.

XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 PT
 XX
 PS Claim 23; Fig 12; 66pp; English.

XX The present sequence is a mutated version of a fusion protein created
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
 CC proteins are both essential for the replication of the virus, acting to
 CC cleave its replicative proteins from the polyprotein produced from the
 CC HCV genome. Inhibitors of the two proteins should be effective as
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A
 CC fusion proteins which can be used to identify inhibitors of this type, as
 CC well as enabling structural studies of the protease and
 CC protease:inhibitor complexes. This sequence contains the alpha-helix0-1
 CC variant.

XX Sequence 195 AA:

Query Match 92.1%; Score 939; DR 21; Length 195;
 Best Local Similarity 93.4%; Pred. No. 9.2e-90;
 Matches 184; Conservative 3; Mismatches 8; Indels 2; Gaps 1;
 QY 1 MKKGGSVVIVGRINLSGDTAYAQOTRGEQGCOKTSHTGRDKNQVEGEVQIVSTATQTFLA 60
 DB 1 MKKGGSVVIVGRIVLNG--AYAQOTRGEQGCQETSQTGRDKNQVEGEVQIVSTAAQTFLA 58
 QY 61 TSINGVLTWVYHGAGTTRTIASPKGPVTOMYTNVDKDLVGMQAPQGSRLTPTCTCGSSDLY 120
 DB 59 TCINGVCWTVYHGAGTTRTIASPKGPVIQMYTNVDKDLVGMQAPQGSRLTPTCTCGSSDLY 118
 QY 121 LVTRHADVIPVRRGDSRGLSPRPISYLGSGGPGLLCPAGHAVGIFRAAVSTRGVAK 180
 DB 119 LVTRHADVIPVRRGDSRGLSPRPISYLGSGGPGLLCPAGHAVGIFRAAVSTRGVAK 178
 QY 181 AVDFIPVESLETTMRSP 197
 DB 179 AVDFIPVESLETTMRSP 195

RESULT 8
 AAB15212
 ID AAB15212 standard; protein: 195 AA.

XX AAB15212;
 AC
 XX
 DT 19-DEC-2000 (first entry)
 XX
 DE Hepatitis C virus NS4A-NS3 fusion protease #1.

XX Hepatitis NS3 protease; viral replication; chronic liver disease;
 KW liver failure; liver cancer.
 KW
 XX Hepatitis C virus.
 OS Synthetic.

XX WO200040707-A1.
 XX
 PD 13-JUL-2000.
 XX
 PF 06-JAN-2000; 2000WO-US00345.
 XX
 PR 08-JAN-1999; 99US-0115271.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

XX
 PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
 XX
 DR WPI: 2000-465976/40.
 DR N-PSDB; AAA73328.

XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 PT
 XX

PS Example 2; Fig 10; 66pp; English.

XX The present sequence is a fusion protein created using the Hepatitis C
 CC virus (HCV) NS3 and NS4A protease enzymes. These proteins are both
 CC essential for the replication of the virus, acting to cleave its
 CC replicative proteins from the polyprotein produced from the HCV genome.
 CC Inhibitors of the two proteins should be effective as antiviral
 CC treatments of HCV infection. This is useful as HCV can lead to chronic
 CC liver disease such as cirrhosis, liver failure and liver cancer. The
 CC present invention concerns a number of NS3 mutants and NS3-NS4A fusion
 CC proteins which can be used to identify inhibitors of this type, as well
 CC as enabling structural studies of the protease and protease:inhibitor
 CC complexes.

XX Sequence 195 AA:

Query Match 89.4%; Score 912; DB 21; Length 195;
 Best Local Similarity 92.4%; Pred. No. 6.1e-87;
 Matches 182; Conservative 1; Mismatches 12; Indels 2; Gaps 1;
 QY 1 MKKGGSVVIVGRINLSGDTAYAQOTRGEQGCOKTSHTGRDKNQVEGEVQIVSTATQTFLA 60
 DB 1 MKKGGSVVIVGRIVLNG--AYAQOTRGLLCITSLTGRDKNQVEGEVQIVSTAAQTFLA 58
 QY 61 TSINGVLTWVYHGAGTTRTIASPKGPVTOMYTNVDKDLVGMQAPQGSRLTPTCTCGSSDLY 120
 DB 59 TCINGVCWTVYHGAGTTRTIASPKGPVIQMYTNVDKDLVGMQAPQGSRLTPTCTCGSSDLY 118
 QY 121 LVTRHADVIPVRRGDSRGLSPRPISYLGSGGPGLLCPAGHAVGIFRAAVSTRGVAK 180
 DB 119 LVTRHADVIPVRRGDSRGLSPRPISYLGSGGPGLLCPAGHAVGIFRAAVSTRGVAK 178
 QY 181 AVDFIPVESLETTMRSP 197
 DB 179 AVDFIPVESLETTMRSP 195

RESULT 9
 AAY24943
 ID AAY24943 standard; protein: 665 AA.

XX AAY24943;
 AC
 XX
 DT 07-SEP-1999 (first entry)
 XX
 DE HCV NS4A-NS3 complex SEQ ID NO:14.

XX HCV; hepatitis C virus; single chain recombinant complex; linker;
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
 KW hydrophobic domain; covalent complex; detection; inhibitor.

XX Hepatitis C virus.
 OS Synthetic.
 XX WO9928482-A2.
 PN
 XX 10-JUN-1999.
 PD
 XX 24-NOV-1998; 98WO-US24528.
 PF
 XX 28-JUL-1998; 98US-0094331.
 PR
 XX 28-NOV-1997; 97US-0067315.

```

XX PA (SCHE ) SCHERING CORP.
XX PI Malcolm BA, Taremi SS, Weber PC, Yao N;
XX DR WPI; 1999-385385/32.
XX PT New hepatitis C virus covalent complexes
XX PS Claim 6; Page 90-92; 21pp; English.
XX CC The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence represents a specifically claimed example of the above
CC complex. The covalent NS4A-NS3 complexes are useful for structural
CC determination and determination of mode of binding of HCV inhibitors by
CC NMR spectroscopy. They can also be used for detecting inhibitors of the
CC protease activity, the helicase activity and the ATPase activity of NS3.
CC The covalent NS4A-NS3 complexes are more soluble, stable and active than
CC the non-covalent protease-peptide complexes previously available.
XX SQ Sequence 665 AA;

Query Match 86.8%; Score 885.5; DB 20; Length 665;
Best Local Similarity 85.7%; Pred. No. 2e-83;
Matches 168; Conservative 15; Mismatches 10; Indels 3; Gaps 1;

QY 5 GSVVIVGRILNSGD---TAYAQOTRGEQGCOKTSHTGRDKNOVEGEVQIVSTATOTFLAT 61
DB 22 GSVVIVGRILNSGSGSTAYSQOTRGLGCKKTSLTGRDKNOVEGEVQIVSTATOSFLAT 81
QY 62 SINGVLMTVTHGAGTRTIA SPKGPVTOMYTNVDKDLVGWQAPGGSRSLTPCTCGSSDLYL 121
DB 82 CVNGVCVTVTHGAGSKTLAGPKGPITQMTYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYL 141
QY 122 VTRHADVIPVRRRGDSRGSLLSPRPISYLKSGSGGPLLCPCAGHAGVGFRAAVSTRGVAKA 181
DB 142 VTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSGGPLLCPSGHAGVGFRAAVCTRGVAKA 201
QY 182 VDFIPVESLETTMRSP 197
DB 202 VDFVPVESMETTMRSP 217

RESULT 10
AAY24947
ID AAY24947 standard; Protein; 665 AA.
XX AC AAY24947;
XX DT 07-SEP-1999 (first entry)
XX DE HCV NS4A-NS3 complex SEQ ID NO:18.
XX KW HCV; hepatitis C virus; single chain recombinant complex; linker;
KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
KW hydrophobic domain; covalent complex; detection; inhibitor.
XX OS Hepatitis C virus.
OS Synthetic.
XX PN WO9928482-A2.
XX PD 10-JUN-1999.
XX PF 24-NOV-1998; 98WO-US24528.
XX PR 28-JUL-1998; 98US-0094331.
PR 28-NOV-1997; 97US-0067315.
XX PA (SCHE ) SCHERING CORP.

```

```

PA (SCHE ) SCHERING CORP.
XX PI Malcolm BA, Taremi SS, Weber PC, Yao N;
XX DR WPI; 1999-385385/32.
XX PT New hepatitis C virus covalent complexes
XX PS Claim 6; Page 100-102; 21pp; English.
XX CC The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence represents a specifically claimed example of the above
CC complex. The covalent NS4A-NS3 complexes are useful for structural
CC determination and determination of mode of binding of HCV inhibitors by
CC NMR spectroscopy. They can also be used for detecting inhibitors of the
CC protease activity, the helicase activity and the ATPase activity of NS3.
CC The covalent NS4A-NS3 complexes are more soluble, stable and active than
CC the non-covalent protease-peptide complexes previously available.
XX SQ Sequence 665 AA;

Query Match 86.5%; Score 882.5; DB 20; Length 665;
Best Local Similarity 85.2%; Pred. No. 4.1e-83;
Matches 167; Conservative 16; Mismatches 10; Indels 3; Gaps 1;

QY 5 GSVVIVGRILNSGD---TAYAQOTRGEQGCOKTSHTGRDKNOVEGEVQIVSTATOTFLAT 61
DB 22 GSVVIVGRILNSGSGSTAYSQOTRGLGCKKTSLTGRDKNOVEGEVQIVSTATOSFLAT 81
QY 62 SINGVLMTVTHGAGTRTIA SPKGPVTOMYTNVDKDLVGWQAPGGSRSLTPCTCGSSDLYL 121
DB 82 CVNGVCVTVTHGAGSKTLAGPKGPITQMTYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYL 141
QY 122 VTRHADVIPVRRRGDSRGSLLSPRPISYLKSGSGGPLLCPCAGHAGVGFRAAVSTRGVAKA 181
DB 142 VTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSGGPLLCPSGHAGVGFRAAVCTRGVAKA 201
QY 182 VDFIPVESLETTMRSP 197
DB 202 VDFVPVESMETTMRSP 217

RESULT 11
AAY24942
ID AAY24942 standard; Protein; 665 AA.
XX AC AAY24942;
XX DT 07-SEP-1999 (first entry)
XX DE HCV NS4A-NS3 complex SEQ ID NO:13.
XX KW HCV; hepatitis C virus; single chain recombinant complex; linker;
KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
KW hydrophobic domain; covalent complex; detection; inhibitor.
XX OS Hepatitis C virus.
OS Synthetic.
XX PN WO9928482-A2.
XX PD 10-JUN-1999.
XX PF 24-NOV-1998; 98WO-US24528.
XX PR 28-JUL-1998; 98US-0094331.
PR 28-NOV-1997; 97US-0067315.
XX PA (SCHE ) SCHERING CORP.

```


XX WPI: 1999-385385/32.
 XX DR New hepatitis C virus covalent complexes
 XX PT Claim 6; Page 97-99; 21lpp; English.
 XX PS The present invention describes a covalent hepatitis C virus (HCV)
 XX CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
 CC to the amino terminus of the HCV NS3 protease domain. The present
 CC sequence represents a specifically claimed example of the above
 CC complex. The covalent NS4A-NS3 complexes are useful for structural
 CC determination and determination of mode of binding of HCV inhibitors by
 CC NMR spectroscopy. They can also be used for detecting inhibitors of the
 CC protease activity, the helicase activity and the ATPase activity of NS3.
 CC The covalent NS4A-NS3 complexes are more soluble, stable and active than
 CC the non-covalent protease-peptide complexes previously available.
 XX Sequence 665 AA;

Query Match 86.1%; Score 878.5; DB 20; Length 665;
 Best Local Similarity 85.2%; Pred. No. 1.1e-82;
 Matches 167; Conservative 15; Mismatches 11; Indels 3; Gaps 1;

QY 5 GSVVIVGRINLSGD---TAYAQOTRGEGCGCKTSHTRDRKNQVEGEVQIVSTATQTFPLAT 61
 DB 22 GSVVIVGRIRIILSGSGSITAYSQOTRGLGCKITSLTGRDRKNQVEGEVQIVSTATQSFPLAT 81
 QY 62 SINGVLWTVYHGAGTRTITASPKGPVTQMTYNVDKDLVGWQAPGQSRSLTPTCTCGSSDLYL 121
 DB 82 CVNGVCWTVYHGAGSKTLAGPKGPITQMTYNVDODLVGWOAPPGARSLTPTCTCGSSDLYL 141
 QY 122 VTRHADVIPVRRRDRSGSLSPRPISYLGSGGGLLCPAGHAGVIFRAAVSTRGVAKA 181
 DB 142 VTRHADVIPVRRRDRSGSLSPRPVSYLKGSGGGLLCPSGHAGVIFRAAVCTRGVAKA 201
 QY 182 VDFIPVESLETTMRSP 197
 DB 202 VDFVPVESMETMRSP 217

RESULT 14
 AAY24941
 ID AAY24941 standard; Protein: 665 AA.
 XX AC AAY24941;
 XX DT 07-SEP-1999 (first entry)
 XX DE HCV NS4A-NS3 complex SEQ ID NO:12.
 XX KW HCV; hepatitis C virus; single chain recombinant complex; linker;
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
 KW hydrophobic domain; covalent complex; detection; inhibitor.
 XX OS Hepatitis C virus.
 OS Synthetic.
 XX PN WO9928482-A2.
 XX PD 10-JUN-1999.
 XX PF 24-NOV-1998; 98WO-US24528.
 XX PR 28-JUL-1998; 98US-0094331.
 XX PR 28-NOV-1997; 97US-0067315.
 XX PA (SCHE) SCHERING CORP.
 XX PI Malcolm BA, Taremi SS, Weber PC, Yao N;
 XX WPI: 1999-385385/32.

DR WPI: 1999-385385/32.
 XX DR New hepatitis C virus covalent complexes
 XX PT Claim 6; Page 85-87; 21lpp; English.
 XX PS The present invention describes a covalent hepatitis C virus (HCV)
 XX CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
 CC to the amino terminus of the HCV NS3 protease domain. The present
 CC sequence represents a specifically claimed example of the above
 CC complex. The covalent NS4A-NS3 complexes are useful for structural
 CC determination and determination of mode of binding of HCV inhibitors by
 CC NMR spectroscopy. They can also be used for detecting inhibitors of the
 CC protease activity, the helicase activity and the ATPase activity of NS3.
 CC The covalent NS4A-NS3 complexes are more soluble, stable and active than
 CC the non-covalent protease-peptide complexes previously available.
 XX Sequence 665 AA;

Query Match 86.0%; Score 877.5; DB 20; Length 665;
 Best Local Similarity 85.2%; Pred. No. 1.4e-82;
 Matches 167; Conservative 15; Mismatches 11; Indels 3; Gaps 1;

QY 5 GSVVIVGRINLSGD---TAYAQOTRGEGCGCKTSHTRDRKNQVEGEVQIVSTATQTFPLAT 61
 DB 22 GSVVIVGRIRIILSGSGSITAYSQOTRGLGCKITSLTGRDRKNQVEGEVQIVSTATQSFPLAT 81
 QY 62 SINGVLWTVYHGAGTRTITASPKGPVTQMTYNVDKDLVGWQAPGQSRSLTPTCTCGSSDLYL 121
 DB 82 CVNGVCWTVYHGAGSKTLAGPKGPITQMTYNVDODLVGWOAPPGARSLTPTCTCGSSDLYL 141
 QY 122 VTRHADVIPVRRRDRSGSLSPRPISYLGSGGGLLCPAGHAGVIFRAAVSTRGVAKA 181
 DB 142 VTRHADVIPVRRRDRSGSLSPRPVSYLKGSGGGLLCPSGHAGVIFRAAVCTRGVAKA 201
 QY 182 VDFIPVESLETTMRSP 197
 DB 202 VDFVPVESMETMRSP 217

RESULT 15
 AAY17884
 ID AAY17884 standard; Protein: 216 AA.
 XX AC AAY17884;
 XX DT 07-SEP-1999 (first entry)
 XX DE HCV NS4A-NS3 complex SEQ ID NO:8.
 XX KW HCV; hepatitis C virus; single chain recombinant complex; linker;
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
 KW hydrophobic domain; covalent complex; detection; inhibitor.
 XX OS Hepatitis C virus.
 OS Synthetic.
 XX PN WO9928482-A2.
 XX PD 10-JUN-1999.
 XX PF 24-NOV-1998; 98WO-US24528.
 XX PR 28-JUL-1998; 98US-0094331.
 XX PR 28-NOV-1997; 97US-0067315.
 XX PA (SCHE) SCHERING CORP.
 XX PI Malcolm BA, Taremi SS, Weber PC, Yao N;
 XX WPI: 1999-385385/32.

A: Variety: isolate JK1

C: Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 23-Mar-2001

C: Accession: S18030; S33570; A48332; S18029

A: Cross-references: EMBL:X61536; NID:g59478; PIDN:CAA43793.1; PID:g59479

A: Experimental source: isolate JK1 from an individual

R: Honda, M.; Kaneo, S.; Masashi, U.; Kobayashi, K.; Murakami, S.

submitted to the EMBL Data Library, September 1991

A: Description: A whole genome of hepatitis C virus CDNA was isolated from a single p.p.

A: Reference number: S18028

A: Accession: S18030

A: Molecule type: genomic RNA

A: Residues: 1-3010 <R0N>

A: Cross-references: EMBL:X61536; NID:g59478; PIDN:CAA43793.1; PID:g59479

A: Experimental source: isolate JK1 from an individual

R: Honda, M.; Kaneo, S.; Unoura, M.; Kobayashi, K.; Murakami, S.

Arch. Virol. 128, 163-169, 1993

A: Title: Sequence analysis of putative structural regions of hepatitis C virus isolate JK1

A: Reference number: A48332; MUID:93119270; PMID:8380322

A: Accession: S33570

A: Molecule type: genomic RNA

A: Residues: 1-547,'T',549-621,'V',623-624,'S',626-652,'DL',655-761,'T',763-782 <H0W>

A: Cross-references: EMBL:X61591

A: Note: this sequence is inconsistent with the nucleotide translation

A: Note: the authors translated the codon AGG for residue 43 as Pro, TGG for residue 45 as Trp, and TTC for residue 771 as Ser

A: Note: sequence extracted from NCBI backbone (NCBIN:121747, NCBIP:121748)

C: Superfamily: hepatitis C virus genome polyprotein

C: Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; se

F: 2-115/Product: capsid protein C #status predicted <CPC>

F: 116-191/Product: envelope protein E #status predicted <EPM>

F: 192-389/Product: major envelope protein E #status predicted <MEE>

F: 390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F: 730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F: 1007-1615/Product: hepatitis C virus polyprotein NS3 #status predicted <NS3>

F: 1230-1237/Region: nucleotide-binding motif A (P-loop)

F: 1312-1317/Region: nucleotide-binding motif B

F: 1316-1319/Region: DEXH motif

F: 1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>

F: 1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>

F: 2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

F: 196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate

Query Match 79.6%; Score 811.5; DB 1; Length 3010;

Best Local Similarity 76.08; Pred. No. 8,7e-66;

Matches 155; Conservative 20; Mismatches 20; Indels 9; Gaps 1;

QY 3 KKGSVIVGRIN-----LSGDTAYAQOTRGCGCKTSHTRGDKNOVEGEVQIVST 53

Db 1005 RRGRELLGADGDFREGQWLLAPITATYISQOTRGLFCIVTSLTRGDKNOVEGEAQVYST 1064

QY 54 ATQTFATLSNGVLWTVYHGAGVTRTASPKGPVTQMTYNDVKDLVWQAPGSSSLPCT 113

Db 1065 ATQSFLATGVGCVTVYHGAGSKTLAGPKPINQMTYNDQDLVWQAPSGAASLPCT 1124

QY 114 CGSSDLYLVTRHADVIPVRRGRSGSLLSPRPISYLYKSGSGGPLLCAGHAGVIFRAAV 173

Db 1125 YGSSDLYLVTRHADVIPVRRGRSGSLLSPRPISYLYKSGSGGPLLCAGHAGVIFRAAV 1184

QY 174 STRGVAKAVDFIPVESLETTMRSP 197

Db 1185 CTGKAVAKAVDFIPVESMETTMRSP 1208

RESULT 9

JC5620

genome polyprotein - hepatitis C virus (isolate EUH1480)

N: contains: capsid protein C; envelope protein M; hepatitis C virus (nonstr

protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C: Species: hepatitis C virus

C: Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001

C: Accession: JC5620

R: Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.

Biochem. Biophys. Res. Commun. 236, 44-49, 1997

A: Title: The complete coding sequence of hepatitis C virus genotype 5a, the predominant

A: Reference number: JC5620; MUID:97366593; PMID:9223423


```

QY 19 TAYAAQTRGEQCGQKTSHTGRDKNOVEGEVOIVSTATQTFELATISNGVLTWYVYHGAGT 78
Db 1034 TAYAAQTRGLLGTIVVSMFGTRDKTRQAGEIQVLTSTVTSFGLTGISGVLTWYVYHGAGNKT 1093

QY 79 IASPKGPVTOMYTNVDKLVGNOAFOGSGSLTPTCGSSDLYLVTRHADVIVPVRRRGDSR 138
Db 1094 LAGSRGPVTQMYSSAEGDLVGPSPGTSKLEPCTCGGAVDLYLVTRNADVIVPARRRGDKR 1153

QY 139 GSLSPRPISYLGKSGGGPLLCAGHAGVIFRAAVSTRGVAKAVDFIPVESLETTMRSP 197
Db 1154 GALLSPRPISLTKSGSGGPPVLCPRGHAGVIFRAAVSTRGVAKAVDFIPVELDIVTMRSP 1212

RESULT 12
T08841
polyprotein - douroucouli hepatitis GB virus A
C:Species: douroucouli hepatitis GB virus A
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000
C:Accession: T08841
R:Erker, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.
J. Gen. Virol. 79, 41-45, 1998
A:Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.
A:Reference number: Z16486; MUID:98120818; PMID:9460920
A:Accession: T08841
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-3005 <ERK>
A:Cross-references: EMBL:AF023425; NID:q2828599; PIDN:AAC40502.1; PID:q2828600
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: polyprotein

Query Match 24.6%; Score 251; DB 2; Length 3005;
Best Local Similarity 34.1%; Pred. No. 2.6e-14;
Matches 56; Conservative 29; Mismatches 69; Indels 10; Gaps 3;

QY 33 KTSHTGRDKNQVEGOIVSTATQTFELATISNGVLTWYVYHGAGTTRIASPKGPVTOMYTN 92
Db 995 KTSMLGRDEREHSJVLGTSTRTSMGTGVNGVMTTFHGSNARTLAGPVGPNCRWS 1054

QY 93 VDKDLVGWQAPGSGSLTPTCGSSDLYLVTRHADVIVPVRRRGDSRSLSPRPISYLGK 152
Db 1055 PSSDVAVYPLPSCAGSCLPECKGTQSWCIRN--DGALCHGRSLKLVLDLPTLSDFRG 1112

QY 153 SSGGPLLCAGHAGVIFRAAVSTRGV-----AKAVDFIPVES 189
Db 1113 SSGSPILCDGHHVGMW-VSVLRGVKVTGVYRKPWETLPKDS 1155

RESULT 13
T08839
polyprotein - marmoset hepatitis GB virus A
C:Species: marmoset hepatitis GB virus A
C>Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 17-Nov-2000
C:Accession: T08839
R:Erker, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.
J. Gen. Virol. 79, 41-45, 1998
A:Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.
A:Reference number: Z16486; MUID:98120818; PMID:9460920
A:Accession: T08839
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: genomic RNA
A:Residues: 1-2970 <ERK>
A:Cross-references: EMBL:AF023424; NID:q2828597; PIDN:AAC40501.1; PID:q2828598
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: polyprotein

Query Match 24.0%; Score 245; DB 2; Length 2970;
Best Local Similarity 27.8%; Pred. No. 9.2e-14;
Matches 62; Conservative 39; Mismatches 80; Indels 42; Gaps 6;

QY 3 KKGSWIVGRIN-----LSGDTAYAAQTRGEQCGQKTSHTGRDKNQVEGEVOIVS 52
Db 946 RRGDEVILIGLVGNWELPFGFVPTAPVYVHHHGKGFYVYKTSMTGWDTEHVGNNVYLG 1005

```

```

QY 53 TATQTFELATISNGVLTWYVYHGAGTTRIASPKGPVTOMYTNVDKLVGNOAFOGSGSLTPTC 112
Db 1006 TSTTRSMGTGVNGVMTTFHGSNARTLAGQMPVNSRWNSADDDVAVIPLPVCAKCLEPC 1065

QY 113 TCGSSDLYLVTRHADVIVPVRRRGDSRGLLS-----PRPISYLGKSGGGPLLCAP 161
Db 1066 KCOPOGVWVI-----RND--GALCHGTLGRVELDLPALCDRFGSGSGSILCD 1112

QY 162 AGHAGVIFRAAVSTRG-----VAKAVDFIPVESLETTMRSP 197
Db 1113 EGHAVGML-ISVLRGSRVTGIRYTKPWETLPREAITHTTEAPP 1154

RESULT 14
H83144
probable aromatic acid decarboxylase PA4019 [imported] - Pseudomonas aeruginosa (str
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: H83144
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; I
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: H83144
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-209 <STO>
A:Cross-references: GB:AE004818; GB:AE004091; NID:q9950200; PIDN:AAG07406.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA4019
C:Superfamily: dedF protein

Query Match 8.4%; Score 85.5; DB 2; Length 209;
Best Local Similarity 27.9%; Pred. No. 1.7;
Matches 51; Conservative 16; Mismatches 61; Indels 55; Gaps 11;

QY 43 QYGEVQ-IVSTATQTFELATISNGVL-----WTVYHGAGTTRIASPKGPVTOMYT 91
Db 29 QREVEVHFLISRAAQLVMTATDVPAKPOAMQAFLEYCGAAGQI-----RVFG 80

QY 92 NYDKDLVGWQAPGSGSLTPTCGSSDL-----YLVTRHADVIVPVRRRGDS 137
Db 81 QND-----WVAPPASGSSAPNAWVPCSTGTLSAVATGACNLIERAADVALKER--- 131

QY 138 RGSLLSPR--PIS-----YLGSSGGPLLCAGHAGVIFRAAVSTRGVAKAVDFIPVES 189
Db 132 RPLVLVPREAPFSSIHLENMLKLSNLGAVILPA--APGFYH---QPOSVELVDVFWVARI 186

QY 190 LET 192
Db 187 LNT 189

RESULT 15
B71284
probable periplasmic serine proteinase DO (htrA-1) - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C>Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 09-Dec-2002
C:Accession: B71284
R:Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; G
rson, J.; Khalak, H.; Richardson, D.; Howell, J.R.; Chidambaram, M.; Utterback, T.; M
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A:Reference number: A71250; MUID:98332770; PMID:9665876
A:Accession: B71284
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-398 <COL>

```

Search completed: August 30, 2003, 19:20:31
Job time : 17.2134 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 18:01:52 : Search time 9.75674 Seconds
(without alignments)
949.524 Million cell updates/sec

Title: US-09-965-594-20

Perfect score: 1020

Sequence: 1 MKKKGSWIVGRINLSGDTA.....VAKAVDFIPVESLETTMRSP 197

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	854.5	83.8	3011	1 POLG_HCV1	P26664 h genome po
2	848.5	83.2	3011	1 POLG_HCVH	P27958 h genome po
3	836.5	82.0	3010	1 POLG_HCVTW	P29846 h genome po
4	827.5	81.1	3010	1 POLG_HCVJT	Q00269 h genome po
5	823.5	80.7	3010	1 POLG_HCVBK	P26663 h genome po
6	823.5	80.7	3010	1 POLG_HCVJA	P26662 h genome po
7	675	66.2	3033	1 POLG_HCVJ8	P26661 h genome po
8	673	66.0	3033	1 POLG_HCVJ6	P26660 h genome po
9	87	8.5	321	1 HHOA_ARATH	Q9se17 arabidopsis
10	85.5	8.4	209	1 PAAD_PSEAE	Q9hx08 pseudomonas
11	83.5	8.2	437	1 DEGI_ARATH	O22609 arabidopsis
12	83	8.1	452	1 AAMP_HUMAN	Q13685 homo sapien
13	78.5	7.7	485	1 Y136_TREPA	O83172 treponema p
14	78.5	7.7	764	1 ICCR_DROME	O08180 drosophila
15	78	7.6	1165	1 POLGALV	P21414 gibbon ape
16	77.5	7.6	263	1 GRAB_MOUSE	O35205 mus musculus
17	76.5	7.5	323	1 VPRT_SMRVH	P21407 squirrel mo
18	76.5	7.5	333	1 MOSA_RHIME	Q07607 rhizobium m
19	76	7.5	401	1 FXHI_MOUSE	O88621 mus musculus
20	76	7.5	3411	1 POLG_YERV1	P03314 y genome po
21	76	7.5	3411	1 POLG_YERV2	P19901 y genome po
22	76	7.5	3414	1 POLG_TBEMV	P14336 t genome po
23	75.5	7.4	248	1 TRY1_CHICK	Q90627 gallus gall
24	75.5	7.4	452	1 MLTD_ECOLI	P23931 escherichia
25	75.5	7.4	2269	1 WDR9_HUMAN	Q9ns16 homo sapien
26	75	7.4	467	1 NX1B_BOVIN	Q28142 bos taurus
27	74.5	7.3	248	1 GRAB_MOUSE	P11033 mus musculus
28	74.5	7.3	3414	1 POLG_LANVT	P29837 l genome po
29	74	7.3	911	1 TB1L_NEITMB	Q05056 neisseria m
30	74	7.3	973	1 VP18_HUMAN	Q9p253 homo sapien
31	74	7.3	3414	1 POLG_TBEMV	Q01299 t genome po
32	73.5	7.2	248	1 TRY2_CHICK	Q90628 gallus gall
33	73.5	7.2	264	1 CTRL_HUMAN	P40313 homo sapien

RESULT 1
POLG_HCV1
ID POLG_HCV1 STANDARD: PRT: 3011 AA.
AC P26664:
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin) DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
DE Hepatitis C virus (isolate 1) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11104;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91172826; PubMed=1848704;
RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C., Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J., Bradley D.W., Kuo G., Houghton M.;
RA "Genetic organization and diversity of the hepatitis C virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION. NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate + (RNA)(N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND M2.
CC PROTEIN C AND M2.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL: M62321; AAA45676.1; -.
CC PIR: A39166; GNVWC3
CC FDB: 1AIV; 16-FEB-99.
CC FDB: 1HEI; 25-NOV-98.
CC MEROPS: S29.001; -.
CC MEROPS: U39.001; -.
CC InterPro: IPR001410; DEAD.
CC InterPro: IPR002522; HCV_capsid.

ALIGNMENTS

34 73.5 7.2 294 1 DPW1_USTMA P54856 ustilago ma
35 73.5 7.2 301 1 MCP_BPF41 P26596 lactococcus
36 73.5 7.2 443 1 FLII_AQUAE O67531 aquifex aeo
37 73.5 7.2 706 1 TRFE_HORSE P27425 equus cabal
38 73.5 7.2 1425 1 NP43_MOUSE P59240 mus musculu
39 73 7.2 478 1 MM03_RABIT P28863 oryctolagus
40 73 7.2 1530 1 NX1A_BOVIN Q28146 bos taurus
41 73 7.2 3412 1 POLG_TBEMV P07720 t genome po
42 73 7.2 3415 1 POLG_POWVL Q04538 t genome po
43 72.5 7.1 660 1 VST2_HEYBU P29326 hepatitis e
44 72.5 7.1 660 1 VST2_HEYBU P33426 hepatitis e
45 72.5 7.1 2499 1 MPRI_BOVIN P08169 bos taurus

DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV env.
 DR InterPro: IPR002531; HCV NS1.
 DR InterPro: IPR002518; HCV NS2.
 DR InterPro: IPR004109; HCV NS3.
 DR InterPro: IPR000745; HCV NS4a.
 DR InterPro: IPR001490; HCV NS4b.
 DR InterPro: IPR002868; HCV NS5a.
 DR InterPro: IPR002166; HCV RdRP.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_Psivir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV env; 1.
 DR Pfam: PF01560; HCV NS1; 1.
 DR Pfam: PF01538; HCV NS2; 1.
 DR Pfam: PF02907; HCV NS3; 1.
 DR Pfam: PF01006; HCV NS4a; 1.
 DR Pfam: PF01001; HCV NS4b; 1.
 DR Pfam: PF01506; HCV NS5a; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00398; Viral_RdRP; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 KW Core protein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 116 191 MATRIX PROTEIN C (POTENTIAL).
 FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4 (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 2014 3011 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 476 476 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;
 Query Match 83.8%; Score 854.5; DB 1; Length 3011;
 Best Local Similarity 82.8%; Pred. No. 2.5e-72;
 Matches 169; Conservative 9; Mismatches 17; Indels 9; Gaps 1;
 QY 3 KGSVVIVGRIN-----LSGDTAYAAQTRGGGCKTSHTGRDNKQVEGEVQIVST 53

Db 1005 RRGREILLGPADGNVSKGMRLLAPITAYAOOTRGLLCITSLTGRDNKQVEGEVQIVST 1064
 QY 54 ATQTFLATSLINGVLTWVYHGAGTRTIASPKGPVTOMYTNVDKDLVGWQAPQGSRLTPTCT 113
 Db 1065 AAOTFLATCLNGVCTWVYHGAGTRTIASPKGPVQMTNTVDQDLVGWPAQGSRLTPTCT 1124
 QY 114 CGSSDLYLVTRHADVIPVRRGDSRGLLSRPISYLGKSSGGPLLCPAGHAGVIFRAAV 173
 Db 1125 CGSSDLYLVTRHADVIPVRRGDSRGLLSRPISYLGKSSGGPLLCPAGHAGVIFRAAV 1184
 QY 174 STGRVAKAVDFIPVESLETTMRSP 197
 Db 1185 CTRGVAKAVDFIPVENLETTMRSP 1208
 RESULT 2
 POLG_HCVH STANDARD; PRT; 3011 AA.
 AC P27958;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate H) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11108;
 RN [1]
 RP MEDLINE=92052256; PubMed=1658800;
 RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
 RA Prince A.M.;
 RT "Genomic structure of the human prototype strain H of hepatitis C
 RT virus: comparison with American and Japanese isolates.";
 Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
 [2]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
 RX MEDLINE=97331322; PubMed=9187654;
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
 "Structure of the hepatitis C virus RNA helicase domain.";
 Nat. Struct. Biol. 4:463-467(1997).
 [3]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
 RX MEDLINE=98154321; PubMed=9493270;
 RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
 RA Murcko M.A., Lin C., Caron P.R.;
 "Hepatitis C virus NS3 RNA helicase domain with a bound
 RT oligonucleotide: the crystal structure provides insights into the mode
 RT of unwinding.";
 Structure 6:89-100(1998).
 RL -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
 CC -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
 CC -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
 CC ACTIVATION OF NS3.
 CC -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
 CC -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC [RNA](N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.

Qy 3 KGSVVIVGRIN-----LSGDTAYAAQOTRGEQGCQKTSHTGRDKNQVEGEVQIVST 53
 Db 1005 RRGREILLGPADSLGRGWRLLAPITAYAAQOTRGLFCIIITSLTGRDKNQVEGEVQIVST 1064
 Qy 54 ATQTFPLATSLVYHAGAGRTTASPKGPTOMYTNWDKDLVGHQAQPGSRSLTPCT 113
 Db 1065 ATOSFLATCLNGVCWTVYHAGAGRTTASPKGPTOMYTNWDKDLVGHQAQPGSRSLTPCT 1124
 Qy 114 CGSSDLYLVTRHADVIVRRGDSRGSLLSPRPISYLGSGGPGLLCPAGHAGVIFRAAV 173
 Db 1125 CGSSDLYLVTRHADVIVRRGDSRGSLLSPRPISYLGSGGPGLLCPAGHAGVIFRAAV 1184
 Qy 174 STRGVAKAVDPIPVSELTMRSP 197
 Db 1185 CTRGVAKAVDPPVSEMETMRSP 1208

RESULT 4

FT POLG_HCVJT STANDARD; PRT; 3010 AA.
 AC Q00269;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein p7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66); (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-JT) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=31642;
 RN [1]
 RP SEQUENCE FROM N.A. PubMed-1318627;
 RA MEDLINE-92295714; PubMed-1318627;
 RA Tanaka T., Kato M., Nakagawa M., Ootsuyama Y., Cho M.J.,
 RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
 RT "Molecular cloning of hepatitis C virus genome from a single Japanese
 RT carrier: sequence variation within the same individual and among
 RT infected individuals."
 RL Virus Res. 23:39-53(1992)
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 CC {RNA}(N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D11168; BAA01943.1; -
 DR PIR: A45573; A45573.
 DR PDB: 1A10; 25-MAR-98.
 DR PDB: 1JXP; 14-JAN-98.
 DR MEROPS: S29.001; -
 DR MEROPS: U39.001; -
 DR InterPro: IPR001410; DEAD.

DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRp.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRp; 1.
 DR Pfam: PD186062; HCV_NS1; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDc; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 1007 1615 NON-STRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 2014 3010 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 347 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT TRANSMEM 347 369 POTENTIAL.
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 3010 AA; 326573 MW; 94A1C77435D0642BB CRC64;

Query Match 81.1%; Score 827.5; DB 1; Length 3010;
 Best Local Similarity 77.0%; Pred. No. 8.9e-70;
 Matches 157; Conservative 20; Mismatches 18; Indels 9; Gaps 1;
 Qy 3 KGSVVIVGRIN-----LSGDTAYAAQOTRGEQGCQKTSHTGRDKNQVEGEVQIVST 53

DB 1005 RRRETLGPDASIEGQWRLLAPITAYAQOTRGLGCVITSLTGRDNQVGEYQVYST 1064
 QY 54 ATQTFLATSLVLTVYHAGTRIASPKGPTVOTMYTNDKDLVGOAPOGSRSLTPTCT 113
 DB 1065 ATQSFATCVGCVTVFHCAGSKTLGPKGPTITMYTNDQDLVGHAPGARSILTPTCT 1124
 QY 114 CGSSDLYLVTRHADYIPVRRRGDSRGLSPRISYLGKSSGGPLLCPCAGHAGVIFRAAV 173
 DB 1125 CGSSDLYLVTRHADYIPVRRRGDSRGLSPRISYLGKSSGGPLLCPCAGHAGVIFRAAV 1184
 QY 174 STRGVAKAVDFIPVESLETTMRSP 197
 DB 1185 CTRGVAKAVDFIPVESMETTMRSP 1208

RESULT 5

POLG_HCVBK STANDARD; PRT: 3010 AA.
 AC P26663;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate BK) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirinae.
 OX NCBI_TaxID=11105;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-91140698; PubMed=1847440;
 RA Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
 RA Onishi E., Andoh T., Yoshida I., Okayama H.;
 RT "Structure and organization of the hepatitis C virus genome isolated
 RT from human carriers.";
 RL J. Virol. 65:1105-1113(1991).
 RN [2]
 RP SEQUENCE OF 1487-1500.
 RX MEDLINE-96235224; PubMed=8647104;
 RA Borowski P., Heiland M., Oehlmann K., Becker B., Kornetevy L.;
 RT "Non-structural protein 3 of hepatitis C virus inhibits
 RT phosphorylation mediated by cAMP-dependent protein kinase.";
 RL Eur. J. Biochem. 237:611-618(1996).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
 RX MEDLINE-97015088; PubMed=8861916;
 RA Love R.A., Farge H.E., Wickersham J.A., Hostomsky Z., Habuka N.,
 RA Moomaw E.W., Adachi T., Hostomsky Z.;
 RT "The crystal structure of hepatitis C virus NS3 proteinase reveals a
 RT trypsin-like fold and a structural zinc binding site.";
 RL Cell 87:331-342(1996).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
 RX MEDLINE-98227846; PubMed=9568891;
 RA Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
 RA Steinkuebler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;
 RT "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C
 RT virus: a 2.2-A resolution structure in a hexagonal crystal form.";
 RL Protein Sci. 7:837-847(1998).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC [RNA](N).

CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M58335; AAA72945.1; -;
 DR PIR; A38465; GNMVTC.
 DR PDB; 1A1Q; 25-MAR-98.
 DR PDB; 1JXP; 14-JAN-98.
 DR PDB; 1NS3; 08-APR-98.
 DR PDB; 1C2P; 15-NOV-00.
 DR PDB; 1GSJ; 08-NOV-99.
 DR PDB; 1GX5; 09-APR-02.
 DR PDB; 1GX6; 10-APR-02.
 DR PDB; 1QTV; 26-JUN-00.
 DR PDB; 80HM; 20-APR-99.
 DR MEROPS; S29.001; -;
 DR
 DR INTERPRO; IPR001410; DEAD.
 DR INTERPRO; IPR002522; HCV capsid.
 DR INTERPRO; IPR002521; HCV core.
 DR INTERPRO; IPR002519; HCV env.
 DR INTERPRO; IPR002531; HCV NS1.
 DR INTERPRO; IPR002518; HCV NS2.
 DR INTERPRO; IPR004109; HCV NS3.
 DR INTERPRO; IPR000745; HCV NS4a.
 DR INTERPRO; IPR001490; HCV NS4b.
 DR INTERPRO; IPR002868; HCV NS5a.
 DR INTERPRO; IPR002166; HCV RdRP.
 DR INTERPRO; IPR007095; RNA_pol_DS_PS.
 DR INTERPRO; IPR007094; RNA_pol_Psvir.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00998; Viral_RdRP; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR PolyProtein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CELLULAR AMINOPEPTIDASE.
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1103
 FT ACT_SITE 1107 1187
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CAPSID PROTEIN C (POTENTIAL).
 FT MATRIX PROTEIN (POTENTIAL).
 FT MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT PROTEASE/HELICASE NS3 (POTENTIAL).
 FT NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT POTENTIAL.
 FT CHARGE RELAY SYSTEM.
 FT CHARGE RELAY SYSTEM.
 FT CHARGE RELAY SYSTEM.
 FT ATP (POTENTIAL).
 FT DECH_BOX.

FT CARBOHYD 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1031 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX 1039 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1050 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1059 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1068 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1075 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1077 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX 1082 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1086 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1090 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1093 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1095 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1101 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1104 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1108 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1120 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1122 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1129 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1135 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1139 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1149 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX 1158 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1162 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1168 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1172 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1175 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1187 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1189 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX 1198 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1203 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1680 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3010 AA; F8422D5ECFCDFD9C CRC64;

Query Match 80.78; Score 823.5; DB 1; Length 3010;
Best Local Similarity 76.5%; Pred. No. 2,4e-69;
Matches 156; Conservative 21; Mismatches 18; Indels 9; Gaps 1;
Qy 3 KKGWVIVGRIN-----LSGDTAYAOOTRGECCQKTSHTGRDKNOVEGEVQIVST 53
Db 1005 RRGKILLGPNADSEGRGLRLAPITAYSQOTRLLGCIITSLTGRDKNOVEGEVQIVST 1064
Qy 54 ATOTFLATSLNGVLWTVYHGAGRTIIASPKGPVTQMTYNDVKULVQWQAPQGGSRSLTPCT 113
Db 1065 ATQSFLATCVNGVCMVTVYHGAGSKTLAAPKGPITQMTYNDVQDQLVGVKPPGARSLLTPCT 1124
Qy 114 CGSSDLXLVTRHADVIPVRRGDSRGSLLSPRPISYILKSGSGGLPLCPAGHANGVIFRAAV 173
Db 1125 CGSSDLXLVTRHADVIPVRRGDSRGSLLSPRPISYILKSGSGGLPLCPAGHANGVIFRAAV 1184
Qy 174 STRGVAKAVDFIPVESLETTMRSP 197
Db 1185 CTRGVAKAVDFIPVESLETTMRSP 1208

RESULT 6
POLG_HCVJA STANDARD; PRT: 3010 AA.
ID POLG_HCVJA
AC P26662;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
(GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
(EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
(EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
NS4B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (HCV).
DE Hepatitis C virus (isolate Japanese) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9108550; PubMed=2175903;
RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
Sugimura I., Shimotohno K.;
RT "Molecular cloning of the human hepatitis C virus genome from
Japanese patients with non-A, non-B hepatitis.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
RN [2]
RP DISCUSSION OF SEQUENCE.
RX MEDLINE=91192160; PubMed=1849488;
RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraio K.,
Ohkoshi S., Shimotohno K.;
RT "Molecular structure of the Japanese hepatitis C viral genome.";
RL FEBS Lett. 280:325-328(1991).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
precursor polyprotein, commonly with Asp or Glu in the P6
position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
[RNA](N).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL: D90208; BAAL4233.1;
DR PIR: A39253; GNMVCGJ.
DR HSSP: P26663; LXP.
DR MEROPS: S29.001;
DR MEROPS: U39.001;
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRP.

DR InterPro: IPR001650; Helicase C.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS4c; 1.
 DR Pfam: PF00271; Helicase C; 1.
 DR Pfam: PF00998; Viral RdRP; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXdc; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 532 532
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2788 2788
 FT SEQUENCE 3010 AA; 327017 MW; AA93794F46DB185 CRC64;
 Query Match 80.7%; Score 823.5; DB 1; Length 3010;
 Best Local Similarity 75.5%; Pred. No. 2.1e-69;
 Matches 154; Conservative 23; Mismatches 18; Indels 9; Gaps 1;
 QY 3 KGSVIVIGRLNSGD-----TAYAQOTRGEQCGCKTSHTGRDKNOVEGQIVST 53
 DB 1005 RRGKEILLGPADSGEQGWRLLAPITAYSQOTRGLLCIITSLTGRDKNOVDGEVQLST 1064
 QY 54 ATQFLATISNGVLWTVYHGAGTFTIAPKGPVTQMTYNDKDLVQWQAPQGSRLTPCT 113
 DB 1065 ATQFLATFCVNGVCWTYVHGAGSTLAPKGPITQMTYNDQDLVGNWAPPAGRSRTPT 1124
 QY 114 CGSSDLYLVTRHADYIPVRRGDSRGLSPRPISYLVKSGSGGLPCPACHAVGIFRAAV 173
 DB 1125 CGSSDLYLVTRHADYIPVRRGDSRGLSPRPISYLVKSGSGGLPCPCHVVGIFRAAV 1184
 QY 174 STRGVAKAVDFIPVESLETTMRSP 197

Db 1185 CTGCAKAVDFIPVESLETTMRSP 1208
 RESULT 7
 POLG_HCVJ8 STANDARD; PRT; 3033 AA.
 AC P26661;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-J8) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11115;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230232; PubMed=1314459;
 RA Okamoto H., Kurai K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,
 RA Fukuda S., Tsuda F., Mishiro S.;
 RT "Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes";
 RL Virology 188:331-341(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC [RNA](N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN C AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND NSNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: D10988; BAA01761.1; -;
 DR PIR: A40250; GNMVJ8.
 DR HSSP: P27958; 1HEI.
 DR MEROPS: S29.001; -;
 DR MEROPS: U39.001; -;
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRP.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.

```

DR Pfam: PF01539; HCV env; 1.
DR Pfam: PF01560; HCV NS1; 1.
DR Pfam: PF01538; HCV NS2; 1.
DR Pfam: PF02907; HCV NS3; 1.
DR Pfam: PF01006; HCV NS4a; 1.
DR Pfam: PF01001; HCV NS4b; 1.
DR Pfam: PF01506; HCV NS5a; 1.
DR Pfam: PF00998; Viral_RdRp; 1.
DR PRODOM: PD186062; HCV NS1; 1.
DR SMART: SM00487; DEXDC; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
FT INIT_MET 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 733
FT CHAIN 734 1010
FT CHAIN 1011 1619
FT CHAIN 1620 1866
FT CHAIN 1867 2017
FT CHAIN 2018 3033
FT TRANSMEM 347 369
FT ACT_SITE 1087 1087
FT ACT_SITE 1111 1111
FT ACT_SITE 1169 1169
FT NP_BIND 1234 1241
FT SITE 1320 1323
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 233 233
FT CARBOHYD 299 299
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 477 477
FT CARBOHYD 534 534
FT CARBOHYD 542 542
FT CARBOHYD 558 558
FT CARBOHYD 578 578
FT CARBOHYD 627 627
FT CARBOHYD 649 649
FT CARBOHYD 1091 1091
FT CARBOHYD 2038 2038
FT CARBOHYD 2359 2359
FT CARBOHYD 2811 2811
SQ SEQUENCE 3033 AA; 330177 MW; 1A173E7E3381FD1A CRC64;

Query Match
Best Local Similarity 66.2%; Score 675; DB 1; Length 3033;
Matches 125; Conservative 24; Mismatches 30; Indels 0; Gaps 0;

QY 19 TAYAOOTRGEQCQKTSHTGRDKNVEGEVQIVSTATOTFLATSIINGVLWTVYHAGTRT 78
DQ 1034 TAYTOOTRGLLGAIVSVLGTGRDKNKGQVQLSSVTTQTLGTSTISGVLWTVYHAGNKT 1093
QY 79 IASPKGPTOMYTKVDKDLGWQAQGSRLTPTCTGSSDLYLVTRHADVIPVRRGDSR 138
DQ 1094 LAGPKGPTQMTYISAEGLDGLWSPSPGKSLDPCPCGAVDLYLVTRNADVIPVRKDDRR 1153
QY 139 GSLLSPRISLYLKGSSGGLPLCPAGHAYGIFRAAYSTRGVAKAVDFIPVESLETTMRSP 197
DQ 1154 GALLSPRLSTLKGSSGGLPCLSRGHAVGLFRAAVCARGVAKSIDFIPVESLDVATRTP 1212

RESULT 8
POLG_HCVJ6
ID POLG_HCVJ6 STANDARD: PRT: 3033 AA.
AC P26660;

```

```

DR Pfam: PF00271: helicase_C; 1.
DR Pfam: PF00998: Viral_RdRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
FT INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 733
FT CHAIN 734 1010
FT CHAIN 1011 1619
FT CHAIN 1620 1866
FT CHAIN 1867 2017
FT CHAIN 2018 3033
FT TRANSMEM 347 369
FT ACT_SITE 1087 1087
FT ACT_SITE 1111 1111
FT ACT_SITE 1169 1169
FT NP_BIND 1234 1241
FT SITE 1320 1323
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 477 477
FT CARBOHYD 534 534
FT CARBOHYD 542 542
FT CARBOHYD 558 558
FT CARBOHYD 578 578
FT CARBOHYD 627 627
FT CARBOHYD 649 649
FT CARBOHYD 1091 1091
FT CARBOHYD 2038 2038
FT CARBOHYD 2811 2811
FT SEQUENCE 3033 AA; 329165 MW; F957F5C1A273BE9E CRC64;

Query Match
Best Local Similarity 66.0%; Score 673; DB 1: Length 3033;
Matches 123; Conservative 27; Mismatches 29; Indels 0; Gaps 0;

QY 19 TAYAOOTRGEQCCQTSHTGRKKNQVEQVQIVSTATOTFLATSIINGVLVTVYHGAGTRT 78
DB 1034 TAYAOOTRGLLGTIVVSMGTGRDKTEQAGEIOVLSTVTSQSLGTTISGLVTVYHGAGNKT 1093
QY 79 IASPKGPVTOMTYNDKDLVGWAOPOGSRSLTPTCCGSSDLYLVRHADYIPVRRRCDNR 138
DB 1094 LAGSRGPTOMYSSREGDLGWSPFGTKSLEPTCCGAVDLYLVRNADYIPARRRDKR 1153
QY 139 GSLLSPRISYLUKSGGGLPLCLPAGHAYGIFRAAVSTRGVAKAYDFIPVRSLETTMRSP 197
DB 1154 GALLSPRLSTLKGSGGPGVLCPRGAVGVFRAAVCSRGVAKSIDFIPVTLDTVTRSP 1212

RESULT 9
HHOA_ARATH STANDARD; PRT; 321 AA.
AC Q9SEL7; O49507;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Protease HhoA, chloroplast precursor (EC 3.4.21.-).
GN HHOA OR AT4G18370 OR F2Bj12.30.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

```

```

OC eutroids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN (1)
RP SEQUENCE FROM N.A.
RA Lensch M.H.A., Sokolenko A., Herrmann R.G.;
RT "Identification and characterization of the chloroplast HhoA protease,
RL a homolog to the bacterial periplasmic protease HhoA."
RN (2)
RP SEQUENCE FROM N.A.
RA Vos P., Hohelsel J., Zimmermann W., Wedler H., Ridley P.,
RA Langham S.A., McCullagh B., Bilham L., Robben J.,
RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
RA Weitzenecker T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA Holzer E., Brandt A., Peters S., van Staveren M., Dirksen W.,
RA Moolijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
RA De Keyser A., Buysshaert C., Gielen J., Villarroel R., De Clercq R.,
RA Van Montagu M., Rogers J., Gronin A., Quail M., Bray-Allen S.,
RA Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
RA Borkova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,
RA Dose S., de Haan M., Maarse A., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fartmann B., Grandrath K., Dauner D., Herzl A.,
RA Neumann S., Argirou A., Vitale D., Liguori R., Pravadini E.,
RA Massenat O., Quigley F., Clabaud L., Muendlein A., Felber R.,
RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA Chedford F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,
RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA Zaccaria P., Devan M., Wilson R.K., de la Bastide M., Habermann K.,
RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA Senkon M., Murray J., Sheet P., Cordes M., Abu-Threiden J.,
RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
RA Antonolu B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S.,
RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA Chen E., Marra M., Martienssen R., McCombie W.R.;
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
thaliana."
RN (3)
RP SEQUENCE OF 72-82; 96-110; 150-159; 178-211 AND 306-320.
RA Schubert M., Peterson U., Funk C., Haas B., Schroeder W.P.,
RA Kieselbach T.;
RT "The chloroplast lumen from Arabidopsis thaliana."
CC 1- SUBCELLULAR LOCATION: Chloroplast; within the thylakoid lumen.
CC 1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
CC 1- CAUTION: Ref.2 sequences differ from that shown due to erroneous
gene model prediction. AT4G18370 and AT4G18375 were originally
fused into a single gene.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its

```

[illegible]

Db 296 GSVDCQAKLYSATTKVGVFRPETAASQPSLGESESNVSLE 341

```
RESULT 13
Y136_TREPA
ID Y136_TREPA STANDARD; PRT: 485 AA.
AC O83172:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical lipoprotein TP0136 precursor.
GN TP0136.
OS Treponema pallidum.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nichols;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete";
RL Science 281:375-388 (1998).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC (Potential).
CC -1- SIMILARITY: BELONGS TO THE TP013X FAMILY OF LIPOPROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE001199; AAC65137.1; ALT_INIT.
DR TIGR: TP0136.
KW Hypothetical protein; Lipoprotein; Membrane; Signal;
KW Complete proteome.
FT SIGNAL 1 23
FT CHAIN 24 485
FT LIPID 24 24
FT LIPID 24 24
FT DOMAIN 164 178
FT DOMAIN 196 210
FT DOMAIN 253 267
FT DOMAIN 318 327
FT DOMAIN 444 447
FT POLY-SER.
FT POLY-SER.
SQ SEQUENCE 485 AA: 48984 MW: C7AACEEDC7DC5CED CRC64;
Query Match 7.7% Score 78.5; DB 1: Length 485;
Best Local Similarity 23.4%, Pred. No. 6.9;
Matches 50; Conservative 16; Mismatches 77; Indels 71; Gaps 10;
QY 16 SGDITAYA-----QOTRGEQCKQTSK-----TGRDKNQVEGVEQIVSTATQTFIATS- 63
Db 54 AGSKLYATNGRLWEKELNGTSGWKVSSSVPTSDK-----KVMSTATDGNIFVLACP 108
QY 64 -NGVLWTVYHGAG-----TFTIASPGQPVQTMVNDKDLVG-----NQAPQGSRLTPCT 113
Db 109 GTGVYKHCVMGAGSSSTGTATSPSTETCSQAT-----LVGGTSPKFLVPGGTGNGNCG 164
QY 114 C-----GSSDLYLVTRHADVP-----VRRGDSRGLSLLSPRISYLK----- 151
Db 165 CGGGGGGSSSSSSCIH1WLVPGGTGNGNCGCGGGGGGSSSSSCIH1KVENTDQFL 224
QY 152 -----GSSGGLLCPAGHVG 167
```

Db 225 DMGEYVVTTKHLYTKNGSSSAGPAQCPCGGGGG 258

```
RESULT 14
ICCR_DROME
ID ICCR_DROME STANDARD; PRT: 764 AA.
AC Q08180;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Irregular chiasm C-roughnest protein precursor (IRREC protein).
GN RST.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=94102535; PubMed=7503814;
RA Ramos R.G., Igloi G.L., Lichte B., Baumann U., Maier D.,
RA Schneider T., Brandstaetter J.H., Froehlich A., Fischbach K.-F.;
RT "The irregular chiasm C-roughnest locus of Drosophila, which affects
RT axonal projections and programmed cell death, encodes a novel
RT immunoglobulin-like protein";
RL Genes Dev. 7:2533-2547 (1993).
CC -1- FUNCTION: REQUIRED FOR CORRECT AXONAL PATHWAY FORMATION IN
CC THE OPTIC LOBE AND FOR PROGRAMMED CELL DEATH IN THE DEVELOPING
CC RETINA.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: POSTEMBRYONIC EXPRESSION IS STRONG IN THE
CC DEVELOPING OPTIC LOBE AND IN THE EYE IMAGINAL DISC.
CC -1- DEVELOPMENTAL STAGE: STRONGLY EXPRESSED IN EMBRYOS. ALSO FOUND
CC IN LATE LARVAL AND PUPAL STAGES.
CC -1- SIMILARITY: Contains 5 immunoglobulin-like C2-type domains
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z21641; CAA79756.1;
CC EMBL: L11040; AAA16632.1;
CC PIR: A49448; A49448.
CC FlyBase: FBgn003285; rst.
CC GO: GO:0016202; P:regulation of myogenesis; IMP.
CC InterPro: IPR007110; Ig-Like.
CC InterPro: IPR003598; Ig-C2.
CC InterPro: IPR003006; Ig_MHC.
CC Pfam: PF00047; Ig; 4.
CC SMART: SM00408; IgC2; 1.
CC PROSITE: PS50835; IG_LIKE; 5.
KW Transmembrane; Immunoglobulin domain; Glycoprotein; Signal; Repeat;
KW Cell adhesion.
FT SIGNAL 1 19
FT CHAIN 20 764
FT DOMAIN 20 533
FT TRANSMEM 534 556
FT DOMAIN 557 764
FT DOMAIN 21 123
FT DOMAIN 117 230
FT DOMAIN 245 261
FT DOMAIN 237 343
FT DOMAIN 346 419
FT DOMAIN 430 530
FT DOMAIN 637 660
FT CARBOHYD 211 211
FT CARBOHYD 313 313
FT CARBOHYD 393 393
FT CARBOHYD 400 400
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 19:00:22 ; Search time 37.5921 Seconds
(without alignments)
1352.314 Million cell updates/sec

Title: US-09-965-594-20

Perfect score: 1020

Sequence: 1 MKRKGSVVIVGRINLSGDTA.....VAKAVDFIPVESLETTMRSP 197

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organalle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	872.5	85.5	4040	12 Q9IFH8	Q9ifh8 mucosal dis
2	858.5	84.2	3011	12 Q36579	Q36579 hepatitis c
3	854.5	83.8	2436	12 Q81756	Q81756 hepatitis c
4	854.5	83.8	3011	12 Q91FE5	Q91fe5 hepatitis c
5	854.5	83.8	3011	12 Q9ELS8	Q9els8 hepatitis c
6	853.5	83.7	3011	12 Q03463	Q03463 hepatitis c
7	851.5	83.5	3011	12 Q36608	Q36608 hepatitis c
8	851.5	83.5	3015	12 Q9PWX5	Q9pwx5 hepatitis c
9	851.5	83.5	3015	12 Q9PWX9	Q9pwx9 hepatitis c
10	849	83.2	181	12 Q91RR8	Q91rr8 hepatitis c
11	849	83.2	181	12 Q91RT5	Q91rt5 hepatitis c
12	847	83.0	181	12 Q91RR5	Q91rr5 hepatitis c
13	847	83.0	181	12 Q91RR2	Q91rr2 hepatitis c
14	847	83.0	181	12 Q91RT9	Q91rt9 hepatitis c
15	846	82.9	181	12 Q91RR3	Q91rr3 hepatitis c
16	846	82.9	181	12 Q91RR4	Q91rr4 hepatitis c

17	846	82.9	181	12 Q91RS1	Q91rs1 hepatitis c
18	846	82.9	181	12 Q91RQ8	Q91rq8 hepatitis c
19	846	82.9	181	12 Q91RT1	Q91rt1 hepatitis c
20	846	82.9	181	12 Q91RR0	Q91rr0 hepatitis c
21	845.5	82.9	3011	12 Q36609	Q36609 hepatitis c
22	844	82.7	181	12 Q91RR6	Q91rr6 hepatitis c
23	844	82.7	181	12 Q91RS9	Q91rs9 hepatitis c
24	843	82.6	181	12 Q91RS3	Q91rs3 hepatitis c
25	842.5	82.6	3011	12 Q9DIT6	Q9dit6 hepatitis c
26	842	82.5	181	12 Q91RT4	Q91rt4 hepatitis c
27	842	82.5	181	12 Q91RS8	Q91rs8 hepatitis c
28	842	82.5	181	12 Q91RT3	Q91rt3 hepatitis c
29	842	82.5	181	12 Q91RS5	Q91rs5 hepatitis c
30	842	82.5	181	12 Q91RS7	Q91rs7 hepatitis c
31	842	82.5	181	12 Q91RT0	Q91rt0 hepatitis c
32	842	82.5	181	12 Q91RS2	Q91rs2 hepatitis c
33	841	82.5	181	12 Q91RS6	Q91rs6 hepatitis c
34	840.5	82.4	3010	12 Q9OP61	Q9op61 hepatitis c
35	840	82.4	181	12 Q91RS4	Q91rs4 hepatitis c
36	839.5	82.3	3010	12 Q68533	Q68533 hepatitis c
37	839	82.3	181	12 Q91RR7	Q91rr7 hepatitis c
38	839	82.3	181	12 Q91RT6	Q91rt6 hepatitis c
39	839	82.3	3011	12 Q36610	Q36610 hepatitis c
40	838	82.2	181	12 Q91RT8	Q91rt8 hepatitis c
41	837.5	82.1	361	12 Q70818	Q70818 hepatitis c
42	837.5	82.1	361	12 Q70817	Q70817 hepatitis c
43	837	82.1	181	12 Q91RR9	Q91rr9 hepatitis c
44	836.5	82.0	3010	12 Q9DTE2	Q9dte2 hepatitis c
45	836.5	82.0	3010	12 Q99AU2	Q99au2 hepatitis c

ALIGNMENTS

RESULT 1

Q9IFH8 ID Q9IFH8 PRELIMINARY; PRT; 4040 AA.
AC Q9IFH8; 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Genome polyprotein.
OS Mucosal disease virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OX Pestivirus.
OC NCBI_TaxID=11099;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20323484; PubMed=10864644;
RA Lai V.C., Zhong W., Skelton A., Ingravallo P., Vassilev V.,
RA Donis R.O., Hong Z., Lau J.Y.;
RT *Generation and characterization of a hepatitis C virus NS3 protease-
RT dependent bovine viral diarrhea virus.*;
RL J. Virol. 74:6339-6347(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Lai V.C.H., Hong Z.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF268278; AAF82566.1; -;
DR HSSP; P26663; 1JXP.
DR MEROPS; S31.001; -;
DR InterPro; IPR000280; CDvir_endptsep80.
DR InterPro; IPR004109; DEAD.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR001650; Helicase.C.
DR InterPro; IPR001005; Myb_DNA_binding.
DR InterPro; IPR001568; RNase_T2.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF00271; Helicase.C; 1.
DR Pfam; PF00998; Viral_RdRP; 1.

DR	Pfam; PF01543; HCV_capsid; 1.
DR	Pfam; PF01542; HCV_core; 1.
DR	Pfam; PF01539; HCV_env; 1.
DR	Pfam; PF01560; HCV_NSI; 1.
DR	Pfam; PF01538; HCV_NS2; 1.
DR	Pfam; PF02907; HCV_NS3; 1.
DR	Pfam; PF01006; HCV_NS4a; 1.
DR	Pfam; PF01001; HCV_NS4b; 1.
DR	Pfam; PF01506; HCV_NS5a; 1.
DR	Pfam; PF00271; helicase_C; 1.
DR	Pfam; PF00998; Viral RdRP; 1.
DR	ProDom; PD186062; HCV_NSI; 1.
DR	SMART; SM00487; DEXDC; 1.
DR	PROSITE; PS50507; RDRP_POSITIVE; 1.
DR	PROSITE; PS50521; RDRP_VIRAL; 1.
KW	ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW	Hydrolase; Nonstructural protein; Polyprotein;
KW	RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ	SEQUENCE 3011 AA; 377182 MW; E2E0E809C63C1B9 CRC64;

Query Match	84.2%; Score 858.5; DB 12; Length 3011;
Best Local Similarity	82.8%; Pred. No. 2.2e-73;
Matches 169; Conservative 10; Mismatches 16; Indels 9; Gaps 1;	

QY	3	KKGSVVIVGRIN;-----LSGDTAYAAQOTRGEQGCOKTSHTGRDKNOYEGEVOIVST	53
DB	1005	RGQEILLGPADGVMVSGWRLLAPIYAAQOTRLLGCIITSLTGRDKNOYEGEVOIVST	1064
QY	54	ATQTFLATISINGLVTVYHGAGTGTITASPKGPVTQMTYNVDKLDYGVQAPQGSRLTPTCT	113
DB	1065	ATQTFLATISINGVCTVYHGAGTGTITASPKGPVIQMTYNVDQDLVGVPAQGSRLTPTCT	1124
QY	114	CGSSDLXLVTRHADVIPVRRGDSRGSLLSPRISYLGSSGGPLLCPAGHAGVIFRAAV	173
DB	1125	CGSSDLXLVTRHADVIPVRRGDSRGSLLSPRISYLGSSGGPLLCPAGHAGVIFRAAV	1184
QY	174	STRGVAKAVDFIPVESLETTMRSP	197
DB	1185	CTRGVAKAVDFIPVENLETTMRSP	1208

RESULT 3			
Q81756			
ID	Q81756	PRELIMINARY;	PRT; 2436 AA.
AC	Q81756;		
DT	01-NOV-1996 (TrEMBLrel. 01, Created)		
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)		
DE	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)		
DE	Genome polyprotein (Fragment).		
OS	Hepatitis C virus.		
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;		
OC	Hepacivirus.		
OX	NCBI_TaxID=11103;		
RA	Choo Q.-L., Richman K., Han J.;		
RA	"The nucleotide sequence of the Hepatitis C viral genome.";		
RL	Submitted (MAY-1990) to the EMBL/GenBank/DBJ databases.		
EMBL	M32084; AAA45677.1; -		
DR	HSP; P27958; 1A1V.		
DR	InterPro; IPR001410; DEAD.		
DR	InterPro; IPR002531; HCV_NSI.		
DR	InterPro; IPR002518; HCV_NS2.		
DR	InterPro; IPR004109; HCV_NS3.		
DR	InterPro; IPR000745; HCV_NS4a.		
DR	InterPro; IPR001490; HCV_NS4b.		
DR	InterPro; IPR002868; HCV_NS5a.		
DR	InterPro; IPR002166; HCV_RdRP.		
DR	InterPro; IPR001650; Helicase_C.		
DR	InterPro; IPR007095; RNA_POL_DS_PS.		
DR	InterPro; IPR007094; RNA_POL_PSVlr.		
Pfam	PF01560; HCV_NSI; 1.		
Pfam	PF01538; HCV_NS2; 1.		

DR	InterPro:	IPR002868;	HCV_NS5a.
DR	InterPro:	IPR002166;	HCV_RdRP.
DR	InterPro:	IPR001650;	Helicase_C.
DR	InterPro:	IPR007095;	RNA_pol_DS_PS.
DR	InterPro:	IPR007094;	RNA_pol_PSVir.
DR	Pfam:	PF01543;	HCV_capsid; 1.
DR	Pfam:	PF01542;	HCV_core; 1.
DR	Pfam:	PF01539;	HCV_env; 1.
DR	Pfam:	PF01560;	HCV_NS1; 1.
DR	Pfam:	PF01538;	HCV_NS2; 1.
DR	Pfam:	PF02907;	HCV_NS3; 1.
DR	Pfam:	PF01006;	HCV_NS4a; 1.
DR	Pfam:	PF01001;	HCV_NS4b; 1.
DR	Pfam:	PF01506;	HCV_NS5a; 1.
DR	Pfam:	PF00271;	helicase_C; 1.
DR	Pfam:	PF00998;	Viral_RdRP; 1.
DR	PRODOM:	PD186062;	HCV_NS1; 1.
DR	SMART:	SM00487;	DEXDC; 1.
DR	PROSITE:	PS00190;	CYTOCHROME_C; 1.
DR	PROSITE:	PS50507;	RDRP_POSITIVE; 1.
DR	PROSITE:	PS50521;	RDRP_VIRAL; 1.
KW	ATP-binding:	Coat protein; Envelope protein; Glycoprotein; Helicase;	
KW	Hydrolase:	Nonstructural protein; Polyprotein;	
KW	kNA-directed RNA polymerase:	Transerase; Transmembrane.	
SO	SEQUENCE	3011 AA; 327124 MW; 2489CE74AC864E58 CRC64;	

Query Match	83.88;	Score 854.5;	DB 12; Length 3011;
Best Local Similarity	82.88;	Pred. No. 5.3e-73;	
Matches 169;	Conservative	9; Mismatches 17; Indels	9; Gaps 1;

Qy	3	KKGSVIVTGRIN-----LSGDTAYAQQTRGEQGCKTSHTGRDKNQVEGEVQIVST	53
	:	: :	:
Db	1005	RGREILLGPADGMVSKGWRLAPITAYAAQTGRLGCIITSLTGDRKNQVEGEVQIVST	1064

Qy	54	ATQTFLATSINGVLWTYYHCAGRTTIASPKGPVTQMNTVDKDLVGHQAPQGSRSLTPTCT	113
	:	: :	:
Db	1065	AATFTLATCINGVCMTYYHGAGRTTIASPKGPVIQMYTNVDQLDVGWPAPQGSRSLTPTCT	1124

Qy	114	CGSSDLYLVTRHADVIPRRRGDSRGSLLSPRPISYLKSGSGGPLLCPAGHVAVFIFRAAV	173
	:	: :	:
Db	1125	CGSSDLYLVTRHADVIPRRRGDSRGSLLSPRPISYLKSGSGGPLLCPAGHVAVFIFRAAV	1184

Qy	174	STRGVAKAVDFIPVESLETTMRSP	197
	:	: :	:
Db	1185	CTRGVAKAVDFIPVENLETTMRSP	1208


```

RESULT 5
Q9ELS8      PRELIMINARY;          PRT:   3011 AA.
AC    Q9ELS8;
DT    01-MAR-2001 (TrEMBLrel. 16, Created)
DD    01-MAR-2003 (TrEMBLrel. 16, Last sequence update)
DT    01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE    Genome polyprotein.
OS    Hepatitis C virus.
OC    Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OX    Hepacivirus.
OA    NCBI_TaxID=11103;
RN    [1]
RC    SEQUENCE FROM N.A.
SC    STRAIN-colonel.
EA    Desai S.M., Devare S., Yamaguchi J.;
RT    "Hepatitis C Virus.";
RL    Submitted (JUL-2000) to the EMBL/GenBank/DBAJ databases.
CC    -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC    LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC    PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC    PROTEIN C AND MRNA (BY SIMILARITY).
DR    EMBL; AF290978; AAG02099.1; -.
DR    HSSP; P27958; 1HEI.
DR    InterPro; IPR000345; CytC_heme_bind.
DR    InterPro; IPR001410; DEAD.

```

```

DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_NS5b.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RDRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS00190; CYTOCHROME_C; 1.
DR PROSITE: PS05057; RDRP_POSITIVE; 1.
DR PROSITE: PS05021; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolyase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA; 327107 MW; A6BECF5A3B3E13F CRC64;

Query Match      83.8%; Score 854.5; DB 12; Length 3011;
Best Local Similarity 82.4%; Pred. No. 5.3e-73;
Matches 168; Conservative 11; Mismatches 16; Indels 9; Gaps 1;

QY 3 KGSWVIVGRIN-----LSGDTAYAQTRGECQCKTSHGRDKNKGVEGVQIVST 53
DB 1005 RGQEILGPADGMVSKGWRLLAPITAYAQTRGLGCIITSLTGRDKNKGVEGVQIVST 1064

QY 54 ATQFLATLSINGLVTVHAGTRTIASPKGPVTOMYTNVDKLVGHQAPQGSRSITPCT 113
DB 1065 ATQFLATCINGVCTVHAGTRTIASPKGPVQIOMTNDQDLVGPAPQGSRSITPCT 1124

QY 114 CGSDDLVLVTRHADVIPRRRGDSRGLSPRPISYLGSGGPGLLCPAGHAYGIFRAAV 173
DB 1125 CGSDDLVLVTRHADVIPRRRGDSRGLSPRPISYLGSGGPGLLCPAGHAYGLFRAAV 1184

QY 174 STRGVAKAVDFIPVESLETMRSP 197
DB 1185 CTRGVAKAVDFIPVENLETMRSP 1208

RESULT 6
Q03463 PRELIMINARY; PRT: 3011 AA.
AC Q03463;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage: Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HC-J1;
RX MEDLINE=91013116; PubMed=2170712;
RA Okamoto H., Okada S., Sugiyama Y., Yotsumoto S., Tanaka T.,
RA Yoshizawa H.;

```

```

RT *The 5'-terminal sequence of the hepatitis C virus genome.*;
RN Jpn. J. Exp. Med. 60:167-177(1990).
RP [2]
RP SEQUENCE FROM N.A.
RC STRAIN-HC-J1;
RX MEDLINE=9204440; PubMed=1658196;
RA Okamoto H., Okada S., Sugiyama Y., Kurai K., Iizuka H., Machida A.,
RA Miyakawa Y., Mayumi M.;
RT *Nucleotide sequences of the genomic RNA of hepatitis C virus isolated
RT from a human carrier: comparison with reported isolates for conserved
RT and divergent regions.*;
RN J. Gen. Virol. 72:2697-2704(1991).
RP [3]
RP SEQUENCE FROM N.A.
RC STRAIN-HC-J1;
RX MEDLINE=93117120; PubMed=1335573;
RA Okamoto H., Kanai N., Mishiro S.;
RT *Full-length nucleotide sequence of a Japanese hepatitis C virus
RT isolate (HC-J1) with high homology to USA isolates.*;
RN Nucleic Acids Res. 20:6410-6410(1992).
RP [4]
RP SEQUENCE FROM N.A.
RC STRAIN-HC-J1;
RA Okamoto H.;
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN-HC-J1;
RX MEDLINE=94174722; PubMed=7510436;
RA Mink M., Benichou S., Madaule P., Tiollais P., Prince A.,
RA Inchausti G.;
RT *Characterization and mapping of a B-cell immunogenic domain in
RT hepatitis C virus E2 glycoprotein using a yeast peptide library.*;
RN Virology 200:246-255(1994).
CC CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA [BY SIMILARITY].
CC EMBL: D10749; BAA01582.1; -.
DR HSP: P27958; 1HEI.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RDRP.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RDRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS05057; RDRP_POSITIVE; 1.
DR PROSITE: PS05021; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolyase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA; 327112 MW; 97E9052C0250463B CRC64;

```

```
Query Match      83.7%; Score 853.5; DB 12; Length 3011;
Best Local Similarity 82.8%; Pred. No. 6.6e-73;
Matches 169; Conservative 8; Mismatches 18; Indels 9; Gaps 1;

QY      3 KKGWVIVGRIN-----LSGDTAYAAQTRGEGCGCKTSHTGRDKNQVEGEVQIVST 53
DB      1005 RRGGEILLGPADGMVSKGWRLLAPITAYAAQTRGLGCIITSLTGRDKNQVEGEVQIVST 1064
QY      54 ATQTFLATSYNGVLVTHYHGAGTRTIA SPKGPVTQMTYNDKDLVGMQAPQGSRLTPTCT 113
DB      1065 AQTFLATCINGVCTVYHGAGTRTIA SPKGPVIOMYTNVDQDLVGMQAPQARSLTPTCT 1124
QY      114 CGSSDLYLVTRHADVIPVRRGDSRGLSPRPISYLGKSSGGPLLCPAGHAGVIFRAAV 173
DB      1125 CGSSDLYLVTRHADVIPVRRGDSRGLSPRPISYLGKSSGGPLLCPAGHAGVIFRAAV 1184
QY      174 STRGVAKAVDFIPVESLETTMRSP 197
DB      1185 CTRGVAKAVDFIPVESLETTMRSP 1208

RESULT 7
O36608          PRELIMINARY:      PRT: 3011 AA.
AC      O36608;
DT      01-JAN-1998 (TREMBLrel. 05, Created)
DT      01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT      01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE      Genome polyprotein.
OS      Hepatitis C virus strain H77.
OC      Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC      Hepacivirus.
OX      NCBI_TaxID=63746;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=H77.
RX      MEDLINE=97385173; PubMed=9238047;
RA      Yanagi M., Purcell R.H., Emerson S.O., Bukh J.;
RT      "Transcripts from a single full-length cDNA clone of hepatitis C virus
RT      are infectious when directly transfected into the liver of a
RT      chimpanzee.";
RL      Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743(1997).
CC      -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC      LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC      PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC      PROTEIN C AND MRNA (BY SIMILARITY).
DB      EMBL; AF011751; AAB67036.1; -
DB      HSSP; P27958; 1HEI
DB      InterPro; IPR001410; DEAD.
DB      InterPro; IPR002522; HCV_capsid.
DB      InterPro; IPR002521; HCV_core.
DB      InterPro; IPR002519; HCV_env.
DB      InterPro; IPR002531; HCV_NS1.
DB      InterPro; IPR002518; HCV_NS2.
DB      InterPro; IPR002519; HCV_NS3.
DB      InterPro; IPR000745; HCV_NS4a.
DB      InterPro; IPR001490; HCV_NS4b.
DB      InterPro; IPR002868; HCV_NS5a.
DB      InterPro; IPR002166; HCV_RdRp.
DB      InterPro; IPR001650; Helicase_C.
DB      InterPro; IPR007095; RNA_pol_DS_PS.
DB      InterPro; IPR007094; RNA_pol_PSVir.
DB      Pfam; PF01543; HCV_capsid; 1.
DB      Pfam; PF01542; HCV_core; 1.
DB      Pfam; PF01539; HCV_env; 1.
DB      Pfam; PF01560; HCV_NS1; 1.
DB      Pfam; PF01538; HCV_NS2; 1.
DB      Pfam; PF02907; HCV_NS3; 1.
DB      Pfam; PF01006; HCV_NS4a; 1.
DB      Pfam; PF01001; HCV_NS4b; 1.
DB      Pfam; PF01506; HCV_NS5a; 1.
DB      Pfam; PF00271; helicase_C; 1.
```

```
DR      Pfam; PF00998; Viral_RdRp; 1.
DR      PRODOM; PD186062; HCV_NS1; 1.
DR      SMART; SM00487; DEXDC; 1.
DR      PROSITE; PS05057; RDRP_POSITIVE; 1.
DR      PROSITE; PS05052; RDRP_VIRAL; 1.
KW      ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW      Hydrolase; Nonstructural protein; Polyprotein;
KW      RNA-directed RNA polymerase; Transferrase; Transmembrane.
SQ      SEQUENCE 3011 AA; 327112 MW; 0B75E681CB5C198 CRC64;

Query Match      83.5%; Score 851.5; DB 12; Length 3011;
Best Local Similarity 82.4%; Pred. No. 1e-72;
Matches 168; Conservative 10; Mismatches 17; Indels 9; Gaps 1;

QY      3 KKGWVIVGRIN-----LSGDTAYAAQTRGEGCGCKTSHTGRDKNQVEGEVQIVST 53
DB      1005 RRGGEILLGPADGMVSKGWRLLAPITAYAAQTRGLGCIITSLTGRDKNQVEGEVQIVST 1064
QY      54 ATQTFLATSYNGVLVTHYHGAGTRTIA SPKGPVTQMTYNDKDLVGMQAPQGSRLTPTCT 113
DB      1065 ATQTFLATCINGVCTVYHGAGTRTIA SPKGPVTQMTYNDQDLVGMQAPQGSRLTPTCT 1124
QY      114 CGSSDLYLVTRHADVIPVRRGDSRGLSPRPISYLGKSSGGPLLCPAGHAGVIFRAAV 173
DB      1125 CGSSDLYLVTRHADVIPVRRGDSRGLSPRPISYLGKSSGGPLLCPAGHAGVIFRAAV 1184
QY      174 STRGVAKAVDFIPVESLETTMRSP 197
DB      1185 CTRGVAKAVDFIPVENLGTMRSP 1208

RESULT 8
Q9PMX5          PRELIMINARY:      PRT: 3015 AA.
AC      Q9PMX5;
DT      01-MAY-2000 (TREMBLrel. 13, Created)
DT      01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT      01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE      Genome polyprotein.
OS      Hepatitis C virus.
OC      Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC      Hepacivirus.
OX      NCBI_TaxID=11103;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      MEDLINE=99420396; PubMed=10489358;
RA      Yanagi M., Purcell R.H., Emerson S.O., Bukh J.;
RT      "Hepatitis C virus: an infectious molecular clone of a second major
RT      genotype (2a) and lack of viability of intertypic 1a and 2a
RT      chimeras.";
RL      Virology 262:250-263(1999).
RN      [2]
RP      SEQUENCE FROM N.A.
RA      Bukh J.;
RL      Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC      -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC      LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC      PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC      PROTEIN C AND MRNA (BY SIMILARITY).
DB      EMBL; AF177040; AAF01182.1; -
DB      EMBL; AF177038; AAF01180.1; -
DB      HSSP; P27958; 1HEI.
DB      InterPro; IPR001410; DEAD.
DB      InterPro; IPR002521; HCV_capsid.
DB      InterPro; IPR002522; HCV_core.
DB      InterPro; IPR002519; HCV_env.
DB      InterPro; IPR002531; HCV_NS1.
DB      InterPro; IPR002518; HCV_NS2.
DB      InterPro; IPR000745; HCV_NS3.
DB      InterPro; IPR001490; HCV_NS4a.
DB      InterPro; IPR001490; HCV_NS4b.
DB      InterPro; IPR002868; HCV_NS5a.
DB      InterPro; IPR002166; HCV_RdRp.
```



```

DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR002129; Pyridoxal dec.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; Helicase_C; 1.
DR Pfam: PF00998; Viral_RDRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS00392; DDC_GAD_HDC_YDC; 1.
DR PROSITE: PS05057; RDRP_POSITIVE; 1.
DR PROSITE: PS05021; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3015 AA; 328159 MW; 87023BCIF190663A CRC64;

Query Match      83.5%; Score 851.5; DB 12; Length 3015;
Best Local Similarity 82.4%; Pred. No. 1e-72;
Matches 168; Conservative 10; Mismatches 17; Indels 9; Gaps 1;

QY 3 KKGWVIVGRIN-----LSGDTAYAAQTREGQCGCKTSHTGRDNKQVEGEVQIVST 53
DB 1009 RGQEILLGPADGMVSKGWRLAPITAYAAQTREGLLGCIITSLTGRDNKQVEGEVQIVST 1068

QY 54 ATQFLATLSINGLVTHYHGAGTRTIASPKGPVTOMTNDKDLVGNQAPQGSRLTPTCT 113
DB 1069 ATQFLATCINGCVTHYHGAGTRTIASPKGPVLOMTNVDODLVGNPAPQGSRLTPTCT 1128

QY 114 CGSSDLVLTTRHADVIPVRRGRSGSLSPRPISYLKSGSGPLLCPCAGHAGVIFRAAV 173
DB 1129 CGSSDLVLTTRHADVIPVRRGRSGSLSPRPISYLKSGSGPLLCPCAGHAGVIFRAAV 1188

QY 174 STRGVAKAVDFIPVESLETMRSP 197
DB 1189 CTRGVAKAVDFIPVENLGTMRSP 1212

RESULT 9
Q9PMU9
ID Q9PMU9 PRELIMINARY; PRT: 3015 AA.
AC Q9PMU9
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_taxid=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99420396; PubMed=10489358;
RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;
RT "Hepatitis C virus: an infectious molecular clone of a second major
RT genotype (2a) and lack of viability of intertypic 1a and 2a
RT chimeras.";
RL Virology 262:250-263(1999).
RP [2]
RP SEQUENCE FROM N.A.
RA Bukh J.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

```

```

CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: AF177039; AAF01181.1; -.
DR EMBL: AF177037; AAF01179.1; -.
DR HSP: P27958; IHEI.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RDRP.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR002129; Pyridoxal dec.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; Helicase_C; 1.
DR Pfam: PF00998; Viral_RDRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS00392; DDC_GAD_HDC_YDC; 1.
DR PROSITE: PS05057; RDRP_POSITIVE; 1.
DR PROSITE: PS05021; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3015 AA; 328084 MW; E309F6318067D6CD CRC64;

Query Match      83.5%; Score 851.5; DB 12; Length 3015;
Best Local Similarity 82.4%; Pred. No. 1e-72;
Matches 168; Conservative 10; Mismatches 17; Indels 9; Gaps 1;

QY 3 KKGWVIVGRIN-----LSGDTAYAAQTREGQCGCKTSHTGRDNKQVEGEVQIVST 53
DB 1009 RGQEILLGPADGMVSKGWRLAPITAYAAQTREGLLGCIITSLTGRDNKQVEGEVQIVST 1068

QY 54 ATQFLATLSINGLVTHYHGAGTRTIASPKGPVTOMTNDKDLVGNQAPQGSRLTPTCT 113
DB 1069 ATQFLATCINGCVTHYHGAGTRTIASPKGPVLOMTNVDODLVGNPAPQGSRLTPTCT 1128

QY 114 CGSSDLVLTTRHADVIPVRRGRSGSLSPRPISYLKSGSGPLLCPCAGHAGVIFRAAV 173
DB 1129 CGSSDLVLTTRHADVIPVRRGRSGSLSPRPISYLKSGSGPLLCPCAGHAGVIFRAAV 1188

QY 174 STRGVAKAVDFIPVESLETMRSP 197
DB 1189 CTRGVAKAVDFIPVENLGTMRSP 1212

RESULT 10
Q91RR8
ID Q91RR8 PRELIMINARY; PRT: 181 AA.
AC Q91RR8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NS3 protease (fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_taxid=11103;

```

```

RN  SEQUENCE FROM N.A.
RP  STRAIN-Pt.1Y;
RA  Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT  "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RL  Clinical Strains of the Hepatitis C Virus.";
DR  EMBL; AF369235; AAK54560.1; -.
DR  InterPro: IPR004109; HCV_NS3.
DR  Pfam: PF02907; HCV_NS3; 1.
KW  Protease.
FT  NON_TER 1 1
FT  NON_TER 181 181
SQ  SEQUENCE 181 AA; 19130 MW; 85D91869299B7C35 CRC64;

Query Match      83.2%; Score 849; DB 12; Length 181;
Best Local Similarity 93.3%; Pred. No. 4.6e-74;
Matches 166; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

QY  19 TAYAAQTREGCGCKTSHTGRDNQVEGEQIVSTATQTFLATSIINGVLTWYVHGAGT 78
DB  4 TAYAAQTRGLLGCIITSLTGRDNQVEGEQIVSTAAQTFLATCINGVCTWYVHGAGT 63
QY  79 IASPKGPVTQMTYNDKDLGWQAPQGSRLTPTCGSSDLYLVTRHADVIPVRRGDSR 138
DB  64 IASPKGPVIQMTYNDKDLGWPAQGSRLTPTCGSSDLYLVTRHADVIPVRRGDSR 123
QY  139 GSLLSPRPISYLGSSGGPLLCAGHAGVIFRAAVSTRGVAKAVDFIPVESLET 196
DB  124 GSLLSPRPISYLGSSGGPLLCAGHAGVIFRAAVSTRGVAKAVDFIPVENLET 181

RESULT 11
Q91RT5
ID  Q91RT5 PRELIMINARY; PRT; 181 AA.
AC  Q91RT5;
DT  01-DEC-2001 (TREMBLrel. 19, Created)
DT  01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT  01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE  NS3 protease (Fragment).
OS  Hepatitis C virus.
OC  Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OX  NCBI_TaxID=11103;
RN  SEQUENCE FROM N.A.
RC  STRAIN-Pt.4;
RA  Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT  "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RL  Clinical Strains of the Hepatitis C Virus.";
DR  EMBL; AF369218; AAK54543.1; -.
DR  InterPro: IPR004109; HCV_NS3.
DR  Pfam: PF02907; HCV_NS3; 1.
KW  Protease.
FT  NON_TER 1 1
FT  NON_TER 181 181
SQ  SEQUENCE 181 AA; 19130 MW; 85D91869299B7C35 CRC64;

Query Match      83.2%; Score 849; DB 12; Length 181;
Best Local Similarity 93.3%; Pred. No. 4.6e-74;
Matches 166; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

QY  19 TAYAAQTREGCGCKTSHTGRDNQVEGEQIVSTATQTFLATSIINGVLTWYVHGAGT 78
DB  4 TAYAAQTRGLLGCIITSLTGRDNQVEGEQIVSTAAQTFLATCINGVCTWYVHGAGT 63
QY  79 IASPKGPVTQMTYNDKDLGWQAPQGSRLTPTCGSSDLYLVTRHADVIPVRRGDSR 138
DB  64 IASPKGPVIQMTYNDKDLGWPAQGSRLTPTCGSSDLYLVTRHADVIPVRRGDSR 123
QY  139 GSLLSPRPISYLGSSGGPLLCAGHAGVIFRAAVSTRGVAKAVDFIPVESLET 196
DB  124 GSLLSPRPISYLGSSGGPLLCAGHAGVIFRAAVSTRGVAKAVDFIPVENLET 181

RESULT 12
Q91RR5
ID  Q91RR5 PRELIMINARY; PRT; 181 AA.
AC  Q91RR5;
DT  01-DEC-2001 (TREMBLrel. 19, Created)
DT  01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT  01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE  NS3 protease (Fragment).
OS  Hepatitis C virus.
OC  Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OX  NCBI_TaxID=11103;
RN  SEQUENCE FROM N.A.
RC  STRAIN-Pt.30;
RA  Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT  "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RL  Clinical Strains of the Hepatitis C Virus.";
DR  EMBL; AF369238; AAK54563.1; -.
DR  InterPro: IPR004109; HCV_NS3.
DR  Pfam: PF02907; HCV_NS3; 1.
KW  Protease.
FT  NON_TER 1 1
FT  NON_TER 181 181
SQ  SEQUENCE 181 AA; 19084 MW; 3B5E8161F2100A72 CRC64;

Query Match      83.0%; Score 847; DB 12; Length 181;
Best Local Similarity 92.7%; Pred. No. 7.2e-74;
Matches 165; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY  19 TAYAAQTREGCGCKTSHTGRDNQVEGEQIVSTATQTFLATSIINGVLTWYVHGAGT 78
DB  4 TAYAAQTRGLLGCIITSLTGRDNQVEGEQIVSTAAQTFLATCINGVCTWYVHGAGT 63
QY  79 IASPKGPVTQMTYNDKDLGWQAPQGSRLTPTCGSSDLYLVTRHADVIPVRRGDSR 138
DB  64 IASPKGPVIQMTYNDKDLGWPAQGSRLTPTCGSSDLYLVTRHADVIPVRRGDSR 123
QY  139 GSLLSPRPISYLGSSGGPLLCAGHAGVIFRAAVSTRGVAKAVDFIPVESLET 196
DB  124 GSLLSPRPISYLGSSGGPLLCAGHAGVIFRAAVSTRGVAKAVDFIPVESLET 181

RESULT 13
Q91RR2
ID  Q91RR2 PRELIMINARY; PRT; 181 AA.
AC  Q91RR2;
DT  01-DEC-2001 (TREMBLrel. 19, Created)
DT  01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT  01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE  NS3 protease (Fragment).
OS  Hepatitis C virus.
OC  Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OX  NCBI_TaxID=11103;
RN  SEQUENCE FROM N.A.
RC  STRAIN-Pt.4V;
RA  Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT  "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RL  Clinical Strains of the Hepatitis C Virus.";
DR  EMBL; AF369241; AAK54566.1; -.
DR  InterPro: IPR004109; HCV_NS3.
DR  Pfam: PF02907; HCV_NS3; 1.
KW  Protease.
FT  NON_TER 1 1
FT  NON_TER 181 181
SQ  SEQUENCE 181 AA; 19123 MW; 1CAE817345ED809D CRC64;
```

```

DB  124 GSLLSPRPISYLGSSGGPLLCAGHAGVIFRAAVSTRGVAKAVDFIPVENLET 181

RESULT 12
Q91RR5
ID  Q91RR5 PRELIMINARY; PRT; 181 AA.
AC  Q91RR5;
DT  01-DEC-2001 (TREMBLrel. 19, Created)
DT  01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT  01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE  NS3 protease (Fragment).
OS  Hepatitis C virus.
OC  Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OX  NCBI_TaxID=11103;
RN  SEQUENCE FROM N.A.
RC  STRAIN-Pt.30;
RA  Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT  "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RL  Clinical Strains of the Hepatitis C Virus.";
DR  EMBL; AF369238; AAK54563.1; -.
DR  InterPro: IPR004109; HCV_NS3.
DR  Pfam: PF02907; HCV_NS3; 1.
KW  Protease.
FT  NON_TER 1 1
FT  NON_TER 181 181
SQ  SEQUENCE 181 AA; 19084 MW; 3B5E8161F2100A72 CRC64;

Query Match      83.0%; Score 847; DB 12; Length 181;
Best Local Similarity 92.7%; Pred. No. 7.2e-74;
Matches 165; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY  19 TAYAAQTREGCGCKTSHTGRDNQVEGEQIVSTATQTFLATSIINGVLTWYVHGAGT 78
DB  4 TAYAAQTRGLLGCIITSLTGRDNQVEGEQIVSTAAQTFLATCINGVCTWYVHGAGT 63
QY  79 IASPKGPVTQMTYNDKDLGWQAPQGSRLTPTCGSSDLYLVTRHADVIPVRRGDSR 138
DB  64 IASPKGPVIQMTYNDKDLGWPAQGSRLTPTCGSSDLYLVTRHADVIPVRRGDSR 123
QY  139 GSLLSPRPISYLGSSGGPLLCAGHAGVIFRAAVSTRGVAKAVDFIPVESLET 196
DB  124 GSLLSPRPISYLGSSGGPLLCAGHAGVIFRAAVSTRGVAKAVDFIPVESLET 181

RESULT 13
Q91RR2
ID  Q91RR2 PRELIMINARY; PRT; 181 AA.
AC  Q91RR2;
DT  01-DEC-2001 (TREMBLrel. 19, Created)
DT  01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT  01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE  NS3 protease (Fragment).
OS  Hepatitis C virus.
OC  Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OX  NCBI_TaxID=11103;
RN  SEQUENCE FROM N.A.
RC  STRAIN-Pt.4V;
RA  Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT  "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RL  Clinical Strains of the Hepatitis C Virus.";
DR  EMBL; AF369241; AAK54566.1; -.
DR  InterPro: IPR004109; HCV_NS3.
DR  Pfam: PF02907; HCV_NS3; 1.
KW  Protease.
FT  NON_TER 1 1
FT  NON_TER 181 181
SQ  SEQUENCE 181 AA; 19123 MW; 1CAE817345ED809D CRC64;
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run On: August 30, 2003, 19:18:33 ; Search time 2560.57 Seconds
(without alignments)
3147.423 Million cell updates/sec

Title: US-09-965-594-20
Perfect score: 1020
Sequence: 1 MKKGSVVIVGRINLSGDTA.....VAKAVDFIPVESLETTMRSP 197

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-O=/cn2_1/vsPTO_spool/US09965594/runat_29082003_151919_28310/app_query.fasta_1.2872
-DB=GenEmbl -OPMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCI=0 -LCOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdl -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09965594.@CGN_1_1.14686.@runat_29082003_151919_28310 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_on:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rtd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_higo_hum:*
40: em_higo_mus:*
41: em_higo_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	892.5	87.5	12734	6	ARI79057	ARI79057 Sequence
2	885.5	86.5	1998	6	ARI45264	ARI45264 Sequence
3	882.5	86.5	1998	6	ARI45268	ARI45268 Sequence
4	881.5	86.4	1998	6	ARI45263	ARI45263 Sequence
5	878.5	86.1	651	6	ARI45254	ARI45254 Sequence
6	878.5	86.1	1998	6	ARI45267	ARI45267 Sequence
7	877.5	86.0	1998	6	ARI45262	ARI45262 Sequence
8	875.5	85.8	651	6	ARI45258	ARI45258 Sequence
9	874.5	85.7	651	6	ARI45253	ARI45253 Sequence
10	874.5	85.7	1998	6	ARI45266	ARI45266 Sequence
11	873.5	85.6	1998	6	ARI45261	ARI45261 Sequence
12	873.5	85.6	2016	6	ARI45269	ARI45269 Sequence
13	872.5	85.5	12734	14	AF268278	AF268278 Pestivirus
14	871.5	85.4	651	6	ARI45257	ARI45257 Sequence
15	870.5	85.3	651	6	ARI45252	ARI45252 Sequence
16	870.5	85.3	1998	6	ARI45265	ARI45265 Sequence
17	870.5	85.3	2016	6	ARI45270	ARI45270 Sequence
18	870	85.3	648	6	ARI45274	ARI45274 Sequence
19	868	85.1	648	6	ARI45272	ARI45272 Sequence
20	867.5	85.0	651	6	ARI45256	ARI45256 Sequence
21	867.5	85.0	651	6	ARI45260	ARI45260 Sequence
22	866.5	85.0	651	6	ARI45251	ARI45251 Sequence
23	866	84.9	648	6	ARI45273	ARI45273 Sequence
24	864	84.7	648	6	ARI45271	ARI45271 Sequence
25	863.5	84.7	651	6	ARI45255	ARI45255 Sequence
26	863.5	84.7	651	6	ARI45259	ARI45259 Sequence
27	861	84.4	8157	6	ARI27810	ARI27810 Sequence
28	861	84.4	8157	6	BD081911	BD081911 Hepatitis
29	859	84.2	1932	6	ARI27809	ARI27809 Sequence
30	859	84.2	1932	6	BD081910	BD081910 Hepatitis
31	858.5	84.2	9646	6	BD110828	BD110828 Sequence
32	858.5	84.2	9646	6	BD069882	BD069882 Functiona
33	858.5	84.2	9646	14	AF009606	AF009606 Hepatitis
34	858.5	84.2	12980	6	ARI10831	ARI10831 Sequence
35	858.5	84.2	12980	6	BD069985	BD069985 Functiona
36	854.5	83.8	5360	6	BD118686	BD118686 Sequence
37	854.5	83.8	5360	6	I06434	I06434 Sequence 48
38	854.5	83.8	5360	6	I09328	I09328 Sequence 8
39	854.5	83.8	6785	6	ARI18692	ARI18692 Sequence
40	854.5	83.8	6785	6	I06440	I06440 Sequence 54
41	854.5	83.8	6785	6	I09329	I09329 Sequence 10
42	854.5	83.8	7310	6	ARI18696	ARI18696 Sequence
43	854.5	83.8	7310	6	I09331	I09331 Sequence 15
44	854.5	83.8	7310	14	HPCPOLYP	M32084 Hepatitis C
45	854.5	83.8	8316	6	ARI18703	ARI18703 Sequence

ALIGNMENTS

```

AR179057
LOCUS AR179057 12734 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6326137.
ACCESSION AR179057
VERSION AR179057.1 GI:20220612
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 12734)
AUTHORS Hong, Z., Lai, V. C. H. and Lau, J. Y. N.
TITLE Hepatitis C virus protease-dependent chimeric pestivirus
JOURNAL Patent: US 6326137-A 1 04-DEC-2001;
FEATURES
Location/Qualifiers
source 1..12734
/organism="unknown"
BASE COUNT 4032 a 2604 c 3295 g 2803 t
ORIGIN
Alignment Scores:
Pred. No.: 15e-65 Length: 12734
Score: 892.50 Matches: 177
Percent Similarity: 92.82% Conservatives: 4
Best Local Similarity: 90.77% Mismatches: 11
Query Match: 87.50% Indels: 3
DB: 6 Gaps: 1
US-09-965-594-20 (1-197) x AR179057 (1-12734)
QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
Db 413 GGTAGTGTGTTATTTGGTAGAATTTATTTATCTGTTAGTGTGTTATCTACAGCGCTAC 472
QY 22 AlaGlnGlnThrArgGlyGluGlnGlyCysGlnIleValSerHisThrGlyValAspLys 41
Db 473 GCGGACAGACGAGAGGCTCTAGGCTGTAAAGATCACCAGTCTGACTGCGCGGACAAA 532
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
Db 533 AACCAAGTGGAGGTGAGCTCCAGATCGTGAACCTGCTACCCAAACCTTCTCGGCAACG 592
QY 62 SerIleAsnGlyValIleuTrpValThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
Db 593 TGCATCAATGGGTATGCTGGACGTCTACCCAGCGCGGACGAGACCATCGCATCA 652
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101
Db 653 CCCAAGGCTCTGTCATCCAGATGTATACCAATGTGGACCAAGACCTTGTGGGTGCC 712
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
Db 713 GCTCTCAAGGTTCCCGTCTATTGACACCCCTGCACCTGCGGCTCTCGGACCTTTAC 772
QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
Db 773 GTTACGAGGACGCGGACGCTATTCCCGTGGCGGCGGAGGTATAGCAGGGGTAGCCTG 832
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
Db 833 CTTTCGCGCGCGCGCATTTCTTACTTAAAGGCTCTCGGGGGTTCGCGTGTGTGCC 892
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
Db 893 GCGGACACGCGCTGGGCTATTACAGCGCGCGGTGTCCACCCGTGGAGTGGCCAAAGCG 952
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196
Db 953 GTGGACTTTATCCCTGTGGAGAACCTAGACAGAACCATGAGATCC 997
RESULT 2
AR145264
LOCUS AR145264 1998 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 105 from patent US 6211338.

```

```

ACCESSION AR145264
VERSION AR145264.1 GI:15107131
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 1998)
AUTHORS Malcolm, B. A., Taremi, S. Shane., Weber, P. C. and Yao, N.
TITLE Single-chain recombinant complexes of hepatitis C virus NS3
protease and NS4A cofactor peptide
JOURNAL Patent: US 6211338-A 105 03-APR-2001;
FEATURES
Location/Qualifiers
source 1..1998
/organism="unknown"
BASE COUNT 411 a 595 c 569 g 423 t
ORIGIN
Alignment Scores:
Pred. No.: 6.96e-66 Length: 1998
Score: 885.50 Matches: 168
Percent Similarity: 93.37% Conservatives: 15
Best Local Similarity: 85.71% Mismatches: 10
Query Match: 86.81% Indels: 3
DB: 6 Gaps: 1
US-09-965-594-20 (1-197) x AR145264 (1-1998)
QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
Db 64 GGTCTGTGTTATTTGGTAGAATTTATTTATCTGTTAGTGTGTTATCTACAGCGCTAC 123
QY 22 AlaGlnGlnThrArgGlyGluGlnGlyCysGlnIleValSerHisThrGlyValAspLys 41
Db 124 TCCCAACAGACGCGGGCTACTTGGTTGCAAGAAGACTAGCTTACAGCGCGGACAAG 183
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
Db 184 AACCAAGTGGAGGTGAGCTCCAGATCGTGAACCTGCTACCCAAACCTTCTCGGCAAC 243
QY 62 SerIleAsnGlyValIleuTrpValThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
Db 244 TGGCTCAAGCGCGCTGTGTGGACCGCTTTACCATGGTGGCTCAAGACCTTAGCCGGC 303
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101
Db 304 CCNAAAGGGCCCAATCACCAGATGTACACTAATGTGGACAGGACCTCGTGGCTGGCAG 363
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
Db 364 GGGCCCCCGGGCGGCTCTTGTACACCATGCACCTGTGGCAGCTCAGACCTTTACTTG 423
QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
Db 424 GTCACGAGACATGCTGACGCTATTCCGCTGCGCGCGGCGGACAGTAGGGGAGCGCTG 483
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
Db 484 CTCCTCCCGCGCGCTCTCTCTACTTGAAGGCTCTTCGGGTGGTCCACTGCTGCGCT 543
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
Db 544 TCGGGGACGCTGTGGGCATCTTCGGGCTGCGGTATGCACCCCGGGGGTTCGGAAGCG 603
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
Db 604 GTGGACTTTGTGCGCGTAGAGTCCATGGAAACTACTATGCGGTCTCCG 651
RESULT 3
AR145268
LOCUS AR145268 1998 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 109 from patent US 6211338.
ACCESSION AR145268
VERSION AR145268.1 GI:15107135

```

KEYWORDS

ORGANISM
Unknown.
Unclassified.

REFERENCE
1 (bases 1 to 1998)

AUTHORS
Malcolm B.A., Taremi S.Shane., Weber P.C. and Yao N.

TITLE
Single-chain recombinant complexes of hepatitis C virus NS3
protease and NS4A cofactor peptide

JOURNAL
Patent: US 6211338-A 109 03-APR-2001;

FEATURES
Location/Qualifiers
source
1..1998

BASE COUNT
411 a 595 c 569 g 423 t

ORIGIN

Alignment Scores:
Pred. No.: 1,25e-65 Length: 1998
Score: 882.50 Matches: 167
Percent Similarity: 93.37% Conservatives: 16
Best Local Similarity: 83.20% Mismatches: 10
Query Match: 86.52% Indels: 3
DB: Gaps: 1

US-09-965-594-20 (1-197) x AR145268 (1-1998)

Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
|||||
Db 64 GGTTCGTGTTATTTGTTAGTAATATTATTTATCTGCTAGTATCATCAGGCGCTAC 123
Qy 22 AlaGlnGlnThrArgGlyGluGlnGlyCysGlnIleValSerThrHisThrGlyArgAspLys 41
|||||
Db 124 TCCACACAGACGGGGGCTACTTGGTTGCAAGAGACTAGCCTTACAGCGCGGAGCAAG 183
Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
|||||
Db 184 AACCAAGTCGAGGAGAGGTTTCAGTGTTCACCGCAACACATCCTTCTGGCGACC 243
Qy 62 SerIleAsnGlyValLeuThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
|||||
Db 244 TGCCTCAACGGCGTGTGTTGGACCGTTTACCATGGTCTGGCTCAAGACCTTAGCGCGC 303
Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101
|||||
Db 304 CCAAGGGGCAATCACCAGATGTACACTAATGTGACCAAGACCTCGTGGCGTGGCAG 363
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
|||||
Db 364 GCGCCCCCGGGGCGCTTCTTGACACCATGACACCTGTGGCAGCTCAGACCTTTACTTG 423
Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
|||||
Db 424 GTCACGAGACATGCTGACGTCAATTCGGGTGCGCGGGCGGCGAGTAGGGGAGCGCTG 483
Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
|||||
Db 484 CTCTCCCCAGCGCTGCTCTCTTCAAGGGCTCTCGGGTGTCTCCAGCTCTGCGCCT 543
Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
|||||
Db 544 TCGGGGACGCTGCGGCACTCTCCGGCTGCCGTATGCACCGGGGGGTGCGAAGCG 603
Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
|||||
Db 604 GTGGACTTTGTGCCGTAGAGTCCATGGAAACTACTATGCGGTCTCGG 651

RESULT 4

AR145263

LOCUS

DEFINITION

Sequence 104 from patent US 6211338.

ACCESSION

AR145263

VERSION

AR145263.1 GI:15107130

KEYWORDS

Unknown.

SOURCE

AR145263

Sequence 104 from patent US 6211338.

ACCESSION

AR145263

VERSION

AR145263.1 GI:15107130

KEYWORDS

Unknown.

SOURCE

AR145263

Sequence 104 from patent US 6211338.

ACCESSION

AR145263

VERSION

AR145263.1 GI:15107130

KEYWORDS

Unknown.

SOURCE

ORGANISM

Unknown.

Unclassified.

REFERENCE

1 (bases 1 to 1998)

AUTHORS

Malcolm B.A., Taremi S.Shane., Weber P.C. and Yao N.

TITLE

Single-chain recombinant complexes of hepatitis C virus NS3

protease and NS4A cofactor peptide

JOURNAL

Patent: US 6211338-A 104 03-APR-2001;

FEATURES

Location/Qualifiers

source

1..1998

BASE COUNT

410 a 596 c 568 g 424 t

ORIGIN

Alignment Scores:

Pred. No.: 1,52e-65 Length: 1998

Score: 881.50 Matches: 168

Percent Similarity: 92.86% Conservatives: 14

Best Local Similarity: 85.71% Mismatches: 11

Query Match: 86.42% Indels: 3

DB: Gaps: 1

US-09-965-594-20 (1-197) x AR145263 (1-1998)

Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
|||||
Db 64 GGTTCGTGTTATTTGTTAGTAATATTATTTATCTGCTAGTATCATCAGGCGCTAC 123
Qy 22 AlaGlnGlnThrArgGlyGluGlnGlyCysGlnIleValSerThrHisThrGlyArgAspLys 41
|||||
Db 124 TCCACACAGACGGGGGCTACTTGGTTGCAAGAGACTAGCCTTACAGCGCGGAGCAAG 183
Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
|||||
Db 184 AACCAAGTCGAGGAGAGGTTTCAGTGTTCACCGCAACACATCCTTCTGGCGACC 243
Qy 62 SerIleAsnGlyValLeuThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
|||||
Db 244 TGCCTCAACGGCGTGTGTTGGACCGTTTACCATGGTCTGGCTCAAGACCTTAGCGCGC 303
Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101
|||||
Db 304 CCAAGGGGCAATCACCAGATGTACACTAATGTGACCAAGACCTCGTGGCGTGGCAG 363
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
|||||
Db 364 GCGCCCCCGGGGCGCTTCTTGACACCATGACACCTGTGGCAGCTCAGACCTTTACTTG 423
Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
|||||
Db 424 GTCACGAGACATGCTGACGTCAATTCGGGTGCGCGGGCGGCGAGTAGGGGAGCGCTG 483
Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
|||||
Db 484 CTCTCCCCAGCGCTGCTCTCTTCAAGGGCTCTCGGGTGTCTCCAGCTCTGCGCCT 543
Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
|||||
Db 544 TCGGGGACGCTGCGGCACTCTCCGGCTGCCGTATGCACCGGGGGGTGCGAAGCG 603
Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
|||||
Db 604 GTGGACTTTGTGCCGTAGAGTCCATGGAAACTACTATGCGGTCTCGG 651

RESULT 5

AR145254

LOCUS

DEFINITION

Sequence 95 from patent US 6211338.

ACCESSION

AR145254

VERSION

AR145254.1 GI:15107121

KEYWORDS

Unknown.

SOURCE

ORGANISM

Unknown.

Unclassified.

AR145254

Sequence 95 from patent US 6211338.

ACCESSION

AR145254

VERSION

AR145254.1 GI:15107121

KEYWORDS

Unknown.

SOURCE

ORGANISM

Unknown.

Unclassified.

AR145254

Sequence 95 from patent US 6211338.

ACCESSION

AR145254

VERSION

AR145254.1 GI:15107121

KEYWORDS

Unknown.

SOURCE

ORGANISM

Unknown.

Unclassified.

REFERENCE 1 (bases 1 to 651)
 AUTHORS Malcolim,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.
 TITLE Single-chain recombinant complexes of hepatitis C virus NS3
 protease and NS4A cofactor peptide
 Patent: US 6211338-A 95 03-APR-2001;

JOURNAL
 FEATURES Location/Qualifiers
 source
 BASE COUNT 120 a 187 c 200 g 144 t
 ORIGIN

Alignment Scores:
 Pred. No.: 7,52e-66 Length: 651
 Score: 878.50 Matches: 167
 Percent Similarity: 93.33% Conservative: 15
 Best Local Similarity: 85.64% Mismatches: 10
 Query Match: 86.13% Indels: 3
 DB: Gaps: 1

US-09-965-594-20 (1-197) x AR145254 (1-651)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
 DB 64 GGTTCGTGTTATCTGTTAGTAATATTTATCTGTTAGTAGTATCAGCGCCTAC 123
 QY 22 AlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThrGlyArgAspLys 41
 DB 124 TCCCAACACACGCGGGCCCTACTTGGTTCGAAGAAGACTAGCTTACAGCGCGGACAAAG 183
 QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
 DB 184 ACCAGGTCGAGGAGAGGTTCAAGTGGTTCCACCGCAACACATCTCTCTGGCGACC 243
 QY 62 SerIleAsnGlyValLeuThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
 DB 244 TGGCTCAACGGCGTGTGGACCGTTTACCATTGGTGGCTCAAGACCTTAGCGGCG 303
 QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
 DB 304 CCAAGGGGCGCAATCACCCAGATGTACACTAATGTGGACCGAGCTGTCGGCTGGCAG 363
 QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
 DB 364 GCGCCCCCGGGCGGTCCTTGACACCATGCACCTGTGGCAGCTCAGACCTTTACTTG 423
 QY 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141
 DB 424 GTCACGAGACATGCTGACGTCATTCGGTGGCGCGGGGCGGACAGTAGGGGAGCGCTG 483
 QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
 DB 484 CTCTCCCCCAGGCGCTGCTCCTACTTGAAGGGCTCTGCTGGTGTCCACTGCTCGCCT 543
 QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
 DB 544 TCGGGGACGCTGTGGGCATCTCCGGGTGCGGTATGCACCCGGGGGTTGCGAAGGCG 603
 QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196
 DB 604 GTGGACTTTGTGCGCGTAGAGTCCATGGAAGTAAGTACTATGCGGTCT 648

RESULT 6
 AR145267

LOCUS AR145267 1998 bp DNA linear PAT 08-AUG-2001
 DEFINITION Sequence 108 from patent US 6211338.
 ACCESSION AR145267
 VERSION AR145267.1 GI:15107134
 KEYWORDS

SOURCE Unknown.
 ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1998)
 AUTHORS Malcolim,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.

TITLE Single-chain recombinant complexes of hepatitis C virus NS3
 protease and NS4A cofactor peptide
 Patent: US 6211338-A 108 03-APR-2001;

JOURNAL
 FEATURES Location/Qualifiers
 source
 BASE COUNT 410 a 596 c 568 g 424 t
 ORIGIN

Alignment Scores:
 Pred. No.: 2,73e-65 Length: 1998
 Score: 878.50 Matches: 167
 Percent Similarity: 92.86% Conservative: 15
 Best Local Similarity: 85.20% Mismatches: 31
 Query Match: 86.13% Indels: 11
 DB: Gaps: 1

US-09-965-594-20 (1-197) x AR145267 (1-1998)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
 DB 64 GGTTCGTGTTATCTGTTAGTAATATTTATCTGTTAGTAGTATCAGCGCCTAC 123
 QY 22 AlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThrGlyArgAspLys 41
 DB 124 TCCCAACACACGCGGGCCCTACTTGGTTCGAAGAAGACTAGCTTACAGCGCGGACAAAG 183
 QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
 DB 184 ACCAGGTCGAGGAGAGGTTCAAGTGGTTCCACCGCAACACATCTCTCTGGCGACC 243
 QY 62 SerIleAsnGlyValLeuThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
 DB 244 TGGCTCAACGGCGTGTGGACCGTTTACCATTGGTGGCTCAAGACCTTAGCGGCG 303
 QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
 DB 304 CCAAGGGGCGCAATCACCCAGATGTACACTAATGTGGACCGAGCTGTCGGCTGGCAG 363
 QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
 DB 364 GCGCCCCCGGGCGGTCCTTGACACCATGCACCTGTGGCAGCTCAGACCTTTACTTG 423
 QY 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141
 DB 424 GTCACGAGACATGCTGACGTCATTCGGTGGCGCGGGGCGGACAGTAGGGGAGCGCTG 483
 QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
 DB 484 CTCTCCCCCAGGCGCTGCTCCTACTTGAAGGGCTCTGCTGGTGTCCACTGCTCGCCT 543
 QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
 DB 544 TCGGGGACGCTGTGGGCATCTCCGGGTGCGGTATGCACCCGGGGGTTGCGAAGGCG 603
 QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
 DB 604 GTGGACTTTGTGCGCGTAGAGTCCATGGAAGTAAGTACTATGCGGTCTCG 651

RESULT 7
 AR145262

LOCUS AR145262 1998 bp DNA linear PAT 08-AUG-2001
 DEFINITION Sequence 103 from patent US 6211338.
 ACCESSION AR145262
 VERSION AR145262.1 GI:15107129
 KEYWORDS

SOURCE Unknown.
 ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1998)
 AUTHORS Malcolim,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.

TITLE Single-chain recombinant complexes of hepatitis C virus NS3
 protease and NS4A cofactor peptide

JOURNAL Patent: US 6211338-A 103 03-APR-2001;

FEATURES

Location/Qualifiers

source

BASE COUNT 410 a 596 c 568 g 424 t

ORIGIN

Alignment Scores:

Pred. No.: 3.33e-65 Length: 1998
Score: 877.50 Matches: 167
Percent Similarity: 92.86% Conservative: 15
Best Local Similarity: 85.20% Mismatches: 11
Query Match: 86.03% Indels: 3
DB: 6 Gaps: 1

US-09-965-594-20 (1-197) x ARI45262 (1-1998)

Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
Db 64 GGTCTGTTGTTATGTTGGTAGAATTATTTATCTGTTAGTGTAGTATCAGCGCTAC 123
Qy 22 AlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThrGlyArgAspLys 41
Db 124 TCCCAACAGACGCGGGCCCTACTGTTGCAAGATCATTAGCTTACAGCGCGGACAG 183
Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
Db 184 AACCAAGTCGAGGAGAGGTTCCAGTGTTTCCACCGCAACAACTCTTCTCGCGACC 243
Qy 62 SerIleAsnGlyValLeuThrValThrValHisGlyAlaGlyThrArgThrIleAlaSer 81
Db 244 TCGCTCAACGGCGTGTGTGACCGTTTACCATGGTCTGGCTCAAGACCTTAGCGCGC 303
Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101
Db 304 CCAAGGGCCAAATCACCAGATGTACATAATGTGACAGGACCTCGTCGGCTGGCAG 363
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
Db 364 GCGCCCGCGGGCGGTCTCTTGCACCATGTCACCTGTGGCAGCTCAGACCTTACTTG 423
Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
Db 424 GTCAGAGACATGCTGACGTCATTCCGGTGGCGGGCGGCGACAGTAGGGGAGCGTG 483
Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
Db 484 CTCTCCCCAGGCGTCTCTCTACTTGAAGGCTCTGCTGGTGTCTGCTGCTGCT 543
Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
Db 544 TCGGGGACGCTGTGGCATCTTCCGGCTGCCGTATGCACCGGGGGGTTGCGAAGCG 603
Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
Db 604 GTGGACTTGTGCGCGTAGAGTCCATGGAAACTACTATGCGGTCTCG 651

RESULT 8

ARI45258

LOCUS

Sequence 99 from patent US 6211338.

ARI45258

DEFINITION

Sequence 99 from patent US 6211338.

ARI45258.1

VERSION

Keywords

Unknown.

ORGANISM

Unknown.

REFERENCE

1 (bases 1 to 651)

AUTHORS

Malcolm,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.

TITLE

Single-chain recombinant complexes of hepatitis C virus NS3

protease and NS4A cofactor peptide

Patent: US 6211338-A 99 03-APR-2001;

Location/Qualifiers

source

BASE COUNT 120 a 187 c 200 g 144 t

ORIGIN

Alignment Scores:

Pred. No.: 1.35e-65 Length: 651
Score: 875.50 Matches: 166
Percent Similarity: 93.33% Conservative: 16
Best Local Similarity: 85.13% Mismatches: 10
Query Match: 85.83% Indels: 3
DB: 6 Gaps: 1

US-09-965-594-20 (1-197) x ARI45258 (1-651)

Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
Db 64 GGTCTGTTGTTATGTTGGTAGAATTATTTATCTGTTAGTGTAGTATCAGCGCTAC 123
Qy 22 AlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThrGlyArgAspLys 41
Db 124 TCCCAACAGACGCGGGCCCTACTGTTGCAAGAGACTAGCTTACAGCGCGGACAG 183
Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
Db 184 AACCAAGTCGAGGAGAGGTTCCAGTGTTTCCACCGCAACAACTCTTCTCGCGACC 243
Qy 62 SerIleAsnGlyValLeuThrValThrValHisGlyAlaGlyThrArgThrIleAlaSer 81
Db 244 TCGCTCAACGGCGTGTGTGACCGTTTACCATGGTCTGGCTCAAGACCTTAGCGCGC 303
Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101
Db 304 CCAAGGGCCAAATCACCAGATGTACATAATGTGACAGGACCTCGTCGGCTGGCAG 363
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
Db 364 GCGCCCGCGGGCGGTCTCTTGCACCATGTCACCTGTGGCAGCTCAGACCTTACTTG 423
Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
Db 424 GTCAGAGACATGCTGACGTCATTCCGGTGGCGGGCGGCGACAGTAGGGGAGCGTG 483
Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
Db 484 CTCTCCCCAGGCGTCTCTCTACTTGAAGGCTCTGCTGGTGTCTGCTGCTGCT 543
Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
Db 544 TCGGGGACGCTGTGGCATCTTCCGGCTGCCGTATGCACCGGGGGGTTGCGAAGCG 603
Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196
Db 604 GTGGACTTGTGCGCGTAGAGTCCATGGAAACTACTATGCGGTCT 648

RESULT 9

ARI45253

LOCUS

Sequence 94 from patent US 6211338.

ARI45253

DEFINITION

Sequence 94 from patent US 6211338.

ARI45253.1

VERSION

Keywords

Unknown.

ORGANISM

Unknown.

REFERENCE

1 (bases 1 to 651)

AUTHORS

Malcolm,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.

TITLE

Single-chain recombinant complexes of hepatitis C virus NS3

protease and NS4A cofactor peptide

Patent: US 6211338-A 94 03-APR-2001;

Location/Qualifiers

source

1. .651

/organism="unknown"

BASE COUNT 119 a 188 c 199 g 145 t
ORIGIN

Alignment Scores:
Pred. No.: 1,64e-65 Length: 651
Score: 874.50 Matches: 167
Percent Similarity: 92.82% Conservative: 14
Best Local Similarity: 85.64% Mismatches: 11
Query Match: 85.74% Indels: 3
DB: 6 Gaps: 1

US-09-965-594-20 (1-197) x ARI45253 (1-651)

Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
Db 64 GGTTCCTGTTATTGTTGGTAGAATATTTATCTGGTAGTGGTAGTATCAGCGCTAC 123
Qy 22 AlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThrGlyArgAspLys 41
Db 124 TCCCAACAGACAGCGGGGCTACTTGGTGCATCAAGACTAGCCTTACAGCGCGGACAG 183
Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
Db 184 AACAGGTGAGGAGAGGTTCAAGTGGTTTCACCGCAACAAATCTTCTGGCGACC 243
Qy 62 SerIleAsnGlyValLeuThrPrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
Db 244 TCGCTCAACGCGGTGCTTGGACCGTTTACCATGGTGTGCTCAAGACCTTAGCGCGC 303
Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
Db 304 CCAAGGGGCAATACCCAGATGTACACTAATGTGGACCAAGACCTCGCGCTGGCAG 363
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
Db 364 GCGCCCCCGGGCGGCTTCTTGACACCATGCACCTGTGGCGAGCTCAGACCTTACTTG 423
Qy 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141
Db 424 GTCAGAGACATGTCAGCTCATTCGGTGGCGGGCGGACAGTAGAGGGAGCGCTG 483
Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
Db 484 CTCTCCCGAGCGCTGCTCTACTTGAAGGGCTCTGCTGGTGGTCCACTGCTGCGCT 543
Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
Db 544 TCGGGGACGCTGTGGGCATCTTCCGGCTGCCGTATGCACCGGGGGGTTCGGAAGCG 603
Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196
Db 604 GTGGACTTGTGGCGGTAGAGTCCATGGAAACTACTATGCGGTCT 648

RESULT 10
ARI45266
LOCUS ARI45266 1998 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 107 from patent US 6211338.
ACCESSION ARI45266
VERSION ARI45266.1 GI:15107133
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 1998)
AUTHORS Malcolim,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.
TITLE Single-chain recombinant complexes of hepatitis C virus NS3 protease and NS4A cofactor peptide
JOURNAL Patent: US 6211338-A 107 03-APR-2001;
FEATURES Location/Qualifiers
source 1..1998
/organism="unknown"

BASE COUNT 410 a 596 c 568 g 424 t
ORIGIN

Alignment Scores:

Pred. No.: 5.98e-65 Length: 1998
Score: 874.50 Matches: 166
Percent Similarity: 92.86% Conservative: 16
Best Local Similarity: 84.69% Mismatches: 11
Query Match: 85.74% Indels: 3
DB: 6 Gaps: 1

US-09-965-594-20 (1-197) x ARI45266 (1-1998)

Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
Db 64 GGTTCCTGTTATTGTTGGTAGAATATTTATCTGGTAGTGGTAGTATCAGCGCTAC 123
Qy 22 AlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThrGlyArgAspLys 41
Db 124 TCCCAACAGACAGCGGGGCTACTTGGTGGCAAGTACACTAGCCTTACAGCGCGGACAG 183
Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
Db 184 AACAGGTGAGGAGAGGTTCAAGTGGTTTCCACCGCAACAAATCTTCTGGCGACC 243
Qy 62 SerIleAsnGlyValLeuThrPrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
Db 244 TCGCTCAACGCGGTGCTTGGACCGTTTACCATGGTGTGCTCAAGACCTTAGCGCGC 303
Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
Db 304 CCAAGGGGCAATACCCAGATGTACACTAATGTGGACCAAGACCTCGCGCTGGCAG 363
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
Db 364 GCGCCCCCGGGCGGCTTCTTGACACCATGCACCTGTGGCGAGCTCAGACCTTACTTG 423
Qy 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141
Db 424 GTCAGAGACATGTCAGCTCATTCGGTGGCGGGCGGACAGTAGAGGGAGCGCTG 483
Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
Db 484 CTCTCCCGAGCGCTGCTCTACTTGAAGGGCTCTGCTGGTGGTCCACTGCTGCGCT 543
Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
Db 544 TCGGGGACGCTGTGGGCATCTTCCGGCTGCCGTATGCACCGGGGGGTTCGGAAGCG 603
Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
Db 604 GTGGACTTGTGGCGGTAGAGTCCATGGAAACTACTATGCGGTCTCG 651

RESULT 11
ARI45261
LOCUS ARI45261 1998 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 102 from patent US 6211338.
ACCESSION ARI45261
VERSION ARI45261.1 GI:15107128
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 1998)
AUTHORS Malcolim,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.
TITLE Single-chain recombinant complexes of hepatitis C virus NS3 protease and NS4A cofactor peptide
JOURNAL Patent: US 6211338-A 102 03-APR-2001;
FEATURES Location/Qualifiers
source 1..1998
/organism="unknown"

BASE COUNT 409 a 597 c 567 g 425 t
ORIGIN

Alignment Scores:

Pred. No.:	7.27e-65	Length:	1998
Score:	873.50	Matches:	167
Percent Similarity:	92.35%	Conservative:	14
Best Local Similarity:	85.20%	Mismatches:	12
Query Match:	85.64%	Indels:	3
DB:	6	Gaps:	1

US-09-965-594-20 (1-197) x AR145261 (1-1998)

Qy	5	GlySerValValIleValIGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr	21
Db	64	GGTCTGTGTTATTGTTGGTAGAATTAATTTATCTGTTAGTGTAGTATCAAGGGCTAC	123
Qy	22	AlaGlnGlnThrArgGlyGluGlnCysGlnLysThrSerHisThrGlyArgAspLys	41
Db	124	TCCCAACAGACGGGGGCTACTTGTTGTCATCATCTAGCTTACAGCCGGGACAAG	183
Qy	42	AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr	61
Db	184	AACCAAGTCAGGGAGAGTTCCAGTGTTTCCACGCCAACACAATCTTCTCGGGACC	243
Qy	62	SerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer	81
Db	244	TGCGTCAACGGGGTGTGTGGACCGTTTACCATGTGTGTGGCTCAAGACCTTAGCGGC	303
Qy	82	ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln	101
Db	304	CCAAAGGGCCAAATCACCACATGTACACTAATGTGGACCAAGGACCTGTCTGGCTGGCAG	363
Qy	102	AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu	121
Db	364	GGCGCCCCGGGGCGGTTCCTTGACACCATGCACCTGTGGCAGCTCAGACCTTACTTG	423
Qy	122	ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu	141
Db	424	GTCACGAGACATGCTGACGCTCATTCGGGTGGCGGGGGGGGCGACAGCTAGGGGGAGCCTG	483
Qy	142	LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro	161
Db	484	CTCTCCCCCGGCCCTGTCTCCCTACTTGAAGGGCTCTTCGGGTGTCTCAGCTCTCGCCCT	543
Qy	162	AlaGlyHisAlaValIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla	181
Db	544	TCGGGGCAGCGTGTGGGCATCTTCGGGCTGCCGTATCCACCCGGGGGTGTGCAGGGCG	603
Qy	182	ValAspPheIleProValIleSerLeuGluThrThrMetArgSerPro	197
Db	604	GTGGACTTGTCCCGTAGTGCATCGAACTACTATGCGGTCTCGC	651

RESULT 12						
AR145269						
LOCUS	AR145269		Sequence	110 bp	DNA	
DEFINITION			Sequence	110 from patent US 6211338.		
ACCESSION	AR145269					
VERSION	AR145269.1					
				GI:15107136		PAT 08-AUG-2001

SOURCE	Unknown.
ORGANISM	Unknown.
	Unclassified.
REFERENCE	1 (bases 1 to 1616)
AUTHORS	Malcolm B.A., Taremi S.Shane., Weber P.C. and Yao,N
TITLE	Single-chain recombinant complexes of hepatitis C virus NS3 protease and NS4A cofactor peptide
JOURNAL	Patent: US 6211338-A 110 03-APR-2001;
FEATURES	Location/Qualifiers
source	1..1616

BASE COUNT	412 a	603 c	570 g	431 t
ORIGIN	/organism-unknown-			

Alignment Scores:	7.34e-65	2016
Pred. No.:	873.50	Length:
Score:		Matches: 167

Percent Similarity:	92.35%	Conservative:	14
Best Local Similarity:	85.20%	Mismatches:	12
Query Match:	85.64%	Indels:	3
DB:	6	Gaps:	1

US-09-965-594-20 (1-197) x AR145269 (1-2016)

Qy	5	GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr	21
Db	82	GGPTCTGTGTTATTGTTGGTAGAATATTATTATCTGTTAGTGTAGTATCACGGCCTAC	143
Qy	22	AlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThrGlyArgAspLys	41
Db	142	TCCCAACACACGGGGGGCCTACTTGTTGCATCATCATCTAGCCTTACAGGCGGGACAAAG	201
Qy	42	AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr	61
Db	202	AACCAGGTGAGGGAGAGGTTTCAGGTGGTTCACCGCCACACACAATCCTCTCGCGACC	261
Qy	62	SerIleAsnGlyValLeuTrpThrValThrHisGlyAlaGlyThrArgThrIleAlaSer	81
Db	262	TGCGTCAACGGCGTGTGTTGGACGGTTTACCATTGCTGCTGCTCAAGACACTTACCGGC	321
Qy	82	ProLysGlyProValThrGlnMetThrThrAsnValAspLysAspLeuValGlyTrpGln	101
Db	322	CCAAAGGGCCCATCACCCAGATGTACACTAATGTGGACACGAGCCTCGTCGGCTGGCAG	381
Qy	102	AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu	121
Db	382	GCSCCCCGGGGGGGTTCCTTGACACCATGCACCTGTGGCAGCTCAGACCTTTACTTG	441
Qy	122	ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu	141
Db	442	GTCAACGACATGCTCAGCTCATTCGGGTGCGCCGGGGCGACAGTGGGGGGAGCGCTG	501
Qy	142	LeuSerProArgProIleSerTyrIleLysGlySerSerGlyGlyProLeuLeuCysPro	161
Db	502	CTCTCCCCAGGCGCTGTCTTACTGTAAAGGGCTCTTCGGGTGGTCCACTGCTCTGCCCT	561
Qy	162	AlaGlyHisAlaValGlyIlePheArgAlaIleValSerThrArgGlyValAlaLysAla	181
Db	562	TCGGGCGACGCTGTGGGCATCTCCGGCTGCCGTATGCACCCGGGGGTTGCCGAGGCG	621
Qy	182	ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro	197
Db	622	GTGACTTTGTGCCCGTAGAGTCCATGGAACTACTATCGCGTCCTCCG	669

RESULT 13
AF268278

LOCUS	AF268278	12734 bp	RNA	linear	VRL 12-JUL-2000			
DEFINITION	Pestivirus type 1, complete genome.							
ACCESSION	AF268278							
VERSION	AF268278.1	GI:9049956						
KEYWORDS								
SOURCE	Pestivirus type 1							
ORGANISM	Pestivirus type 1							
	Viruses; ssRNA positive-strand viruses, no DNA stage: Flaviviridae; Pestivirus.							

REFERENCE
AUTHORS
TITLE
JOURNAL
20232484
MEDLINE
BITMFO
10854644

2 (bases 1 to 12734)
Lai, V.C.H. and Hong, Z.

TITLE Direct Submission
JOURNAL Submitted (16-May-2000) Antiviral Therapy, Schering-Plough Research Institute, 2015 Galloping Hill Road, Kenilworth, NJ 07033-0539, USA
FEATURES Location/Qualifiers
source 1. .12734

```
/organism="Pestivirus type 1"
/mol_type="genomic RNA"
/db_xref="taxon:11099"
1..385
386..12508
/codon_start=1
/product="polyprotein"
/protein_id="AAF82566.1"
/db_xref="GI:9049957"
/translation="MEINTNEGSGSVIVIRGVLGSGSITACAOOTRGLGCKLTSL
TGRKNQVEGEIVSTATOTIFLATCINGVCHTVYHGAGTRTASPKGVIOYINVD
ODLVGWPAPGSHSLPCTCGSDSLVTRHANVIVRRGDSRLSPRISLISLKG
SSGPLCPAGHAGVFLAAVCTRGVAKAVDFIPVENLETTTRSSGADTEDVCCSM
SYDTEEGATKKTOKPDLERKMKIVPKSEKSKTKPPDATTIVGVGVQYRK
GKYSKNTDGLYHNKPKOESRKLEKALWAIIVLVFTVMTGENTONLQDNG
TEGIRAMFORGVNRLHGWPEKICTGVPSHLATDIELKTIGHMDASEKYNCCR
LORHWNKHCWKNYNIPEHILVNRTOANTLEGQPPRECAVTCRVDRAEDLNVTQA
RDSPTLTKGCKNFSFAGILMRPCNFEIAASDVLFKEHERISMFOITLILVDGL
TNSLEGARQGTAKITLWLGKLGILGKLEKSKTWFGAYAAAPYCDVDKIGIYWT
KNTACPLKNTKIVGPFDTNAEDGKILHENGHLSEVLLSLVLSDFEAPETASV
MYLIFPESIQSHVDVMDCKTQNLNLTVELTADVIPGSMVNLGWKVCIRPNWMPYET
TVLAFEEVSQVVKLVLRDLRTIWNAAATTTAFILCLVKTIVRGOMVOSILWLLIT
GVGHLDCKPEFSYAKDRIGOLGAEGLTITWKEYSPOKLEDTWVLANCEDKLM
YLCQRTRETRIALHTRALPTSVYFKKLFDRGRKEDVVMNDNFEFGCLPCDAPIV
RGKFNITLLNGPAPVPCPVGMTGVTSCTSFNMDTLATIVRTYRSKFFPHRQGCIT
QKNGEDLHNCILGNNMTCVPGDQLLYKGSIESCWKCGYQFKESGLPHYPIGRCKL
ENETGYRLDSTSCNREGVAVPQGTILCKIKGTQVQVAMDTKLGPMPCRPYETIIS
EGPVKTAQFNTKTLLKRYFEPDSYFOQYMKGEYQVWPLEVTDHHRDYFAESI
LVVVALLGGRYVLMVLLTVYVLSQKALGQYGVSEVMMGLPHNNIEVTVFLL
LYLLLEESYKVKVLLYHLYVHPISKVITVLLMGDVVKADSGGOEYLGKIDLCFT
TVYLVIGLIIARDDPTIPLVIMAAIRVETLTHQPGVDIAVAVITILLWYSYVD
YFRKMLQCLISLVSQVFLIRSLIYLRITEMPEVTPINWRPULILLIYLISTVTR
WKVDVAGLQCPVILLVTLWADFLILILPLVELKLYLKTVRTDIERSLUGG
IDTRVDSYDDESGBGVTFPPSRQAQNFSLILPLIKATILCSVSKWOLYMSY
LTDIFYMYHRKVYIETSGNTNISRLVAALIELNMBEESKLEKLYLLSGRLRN
LIIKHVNTETASVGEERYGNPKLMTIIRKASTLSKSRHCIIICVCGEWMKGTIC
PKGRCHKPTECGSLADFERHYKRIEIRGNEGMCSCRCOGKRFEMDEPESAR
YCARLHPAEGEDFWAESMLGLKITYFALMDGKVYDITENAGCQRGVISPDHRY
PCHISFOSRPFQOYFVOYTARGOLFURNLPVLATIKVLMNLVNGNLEIGLIDELH
GWLIRGPAKCKITEHEKHINILDKLTAFFGIRPGTTPRAPREPTSLKYRRGLE
TGWAYTHQGGISYDHYTAGKLDLVDSCMGRTRVQSNRNLDETEYGVKTDGSGPD
GARYLVPAVNVKNSKGVAKVHLQGTGCTVAGTAPFDFLKLWGLWSGLPFE
LPLDITVIDRGLCKEKRVYSSKIPETVGLKRMVTVGEQARRGRVGRVPGRYR
SQETATSGDYHVDLQAOQYGEDINVTKSEMNMDLSYEEDSLAITOLEILNN
LLISEDUPAAVKNIMARTDHPETOLANSYEVQVPLFPKIRNGEVTDTYENYSPLN
ALKIGEDVPVYIATEDEDLAVDLGLDWDPCNQOVETGKALKQVOTGLSSAENALL
VALGYVQALSKRHPMTDIYTTEDORLEDTTHLOYAPNAIKDTGETELKELAS
GDVEKIMGALSDTAGGCEVFOAEKIKTAPLEKNAEAAKGVOKFIIDSLKENKEE
IIRYGLGTHALYKSLAARLGHETAFATLVKLAPGGESVDHYKQAADLVYVYV
MKNPSPGDESETQOGERFRFVASLFIISALATYTYTKWNLHNSKVPEPALAYPIATSA
LKMFETPRLESVILSTYIKYLSIRKSGKDLGATGISAAMEILSONPVSVIGSVI
LGVCIAAHNAIESSEKTYLKVRFVKNLDOAADTDELKVENPKIIMALEFAVOTI
GNPLRLIYHLGYVYKWEAKELESETPAGNLPFLIMFEAFELIHDWSOCKIRNLSON
YILDLIYGLHQINRGLKVKVGNAPAPPSCDWTPSDERIRLPTDMLRVTETPCQGY
EMKAFKNGGKTRKVESGFFLCNRRPGRPVRYTKYDDNREIKRIPVAKLEGOVE
HYRGVATKIDYVSKGKMLATDKWEHGVITBLAKRYTGVFGNAYLGDENPHRALY
ERDCATITKNTVQPLKMKKCAFTYDILTISNLTRELHVRNLEKEIPTATPWTWL
VLEKVPDSDNONSVAIGLDGNYPGGLOTHLQTEIHNRPARPIMILGSRNSIS
NRKATARNILYTDNDPREIRDLMAAGRLVVALRDVDPSEKWFKGTFLDREALE
ALSUGQPKQVOTKAVRNLEOKKVEIPNWFASDDPVELEVALKNTYILVDGGE
VKDAQALGATDQTRIIIEKVSRTYAKLSWFLQASNKOMSLTPIFELLRLCPAT
KSNRGNHASATYLAQGWPELGGCVHGTIPARRVKIHPYEAYLKLDKDFEERKPR
VKDTPVIRHNKWLKIRFOGNLNTKMLNPGKLSQOLDREGRKRNLYNHOICTIMSS
AGIRLEKLPVIRACTDITKTHEAIRDKIDKSENQNPENHKLLEIFPHIAQTLKHT
YGEVTEQLEAGINRGAAGFLEKKNIGEVDSEKHLVEOLVRDLKAGARKIYETAI
PKNEKRDYSDMQAGDLVIEKRRPVIQIYPEAKTLTAITKVMVNWVQOQVPIPGYEGK
TPLFNIFDKVRKEWDSNEPVAVSFDTKAWDQTQVTSKDLQILGEIOIKYIKYKWEHKEFI
```

```
DTITDHMTVEVPITADGEVIRMGQSGQPDTSAGNSMLNVLMTMMYAFCESTGVYK
SEFNAVRHVGDDGDFLITEKGLGKFAKNGMOILHEAGKPOKITSEKMKVAYREFD
IEFSHTPVPVWSDNTSSHAGRDPAVILSKMATRLDSSGGERGTIVKAKVAFSEFL
MISWRPLVRRICLUVSQDDEPDKSHATYIKGDPKATGKDVIGVIRNLSKATPEK
LANLNSLSLTGLIWTHTSKRIIODOCAIKGEGNMLVNADRLSLHLYIPDKGF
TLOGKHYYEQQLRTETPNVGVGTGRTYKLGPIVNLRLKILLMTAVGVSS"
12509..12734
BASE COUNT 4030 a 2608 c 3293 g 2802 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 7,45e-64 Length: 12734
Score: 872.50 Matches: 174
Percent Similarity: 91.79% Conservative: 5
Best Local Similarity: 89.23% Mismatches: 13
Query Match: 85.54% Indels: 3
DB: 14 Gaps: 1
US-09-965-594-20 (1-197) x AF268278 (1-12734)
QY 5 GlySerValValIleValGlyValGlyLeuSerGlyAsp-----ThrAlaTyr 21
DB 413 GGTAGTGTGTATTGTTGGTAGAATGTTTATCTGCTAGTGTATCATCGGGTGC 472
QY 22 AlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThrGlyArgAspLys 41
DB 473 GCCCAGCAGCAGCAGAGGCTCTAGGGGTAAAGATCACCAGTCTGAGTGGCGGACAAA 532
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
DB 533 AACCAAGTGGAGGGTGAGGTCCAGATCGTGAACCTGCTACCAACCTTCTTGGCNAAG 592
QY 62 SerIleAsnGlyValLeuTyrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
DB 593 TGCATCAATGGGTATGCTGACGATGCTTACCACGGGCGGAAACGAGACCATCGCATCA 652
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101
DB 653 CCCAAGGTCTCTGTCATCCAGATGATATACCAATGTGACCAAGACCTTGTGGGCGCC 712
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
DB 713 GCTCTCAAGGTTCGGCTCATTCACCCCTGCACCTCGGCTCTCTGGACCTTTACCTG 772
QY 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141
DB 773 GTTAGAGGCGACGCGCAACGATCTCCCGTGGCGGCGAGGTAGTAGCGGGTAGCGCTG 832
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerGlySerGlyProLeuLeuCysPro 161
DB 833 CTTTCGCCCGCGCCCATTTCTACCTACCAAGGCTCCTCTGGGGCTCGCTGTGTGCCCC 892
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
DB 893 GCGGGACACGCGGTGGGCTATTTCAGGGCGCGGTGTGTGACCCGCTGGAGTGGCCAGGCG 952
QY 182 ValAspPheIleProValGluSerLeuGlnThrThrMetArgSer 196
DB 953 GTGGACTTTATCCCTGTGGAGAACCTAGACAGAACCAACGAGATCC 997
RESULT 14
LOCUS AR145257
DEFINITION Sequence 98 from patent US 6211338.
ACCESSION AR145257
VERSION AR145257.1 GI:15107124
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 651)
AUTHORS MalcolM.B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.
TITLE Single-chain recombinant complexes of hepatitis C virus NS3
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 30, 2003, 19:13:57 ; Search time 182.939 Seconds
(without alignments)
2906.924 Million cell updates/sec

Title: US-09-965-594-20

Perfect score: 1020

Sequence: 1 MKKKGWVIVGRINLSGDFA.....VAKAVDFIPVESLETMRSP 197

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-O=/cnp2_1/USPTO.spool/US09965594/runat_29082003_151918_28302/app_query.fasta_1.2872
-DB=N_Geneseq19Jun03 -CFMT=fastp -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cd1
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09965594 -CGN_1_1_1412 -runat_29082003_151918_28302 -NCPU=6 -ICPU=3
-NO_MMAPP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq19Jun03:**
1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:**
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:**
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:**
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:**
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:**
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:**
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:**
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:**
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:**
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:**
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:**
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:**
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:**
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:**
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:**
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:**
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:**
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:**
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:**
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:**
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:**
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:**
23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:**
24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:**
25: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1020	100.0	594	21	AAA73333	Hepatitis C virus
2	1010	99.0	594	21	AAA73334	Hepatitis C virus
3	1005	98.5	594	21	AAA73332	Hepatitis C virus
4	990	97.1	594	21	AAA73331	Hepatitis C virus
5	973	95.4	594	21	AAA73330	Hepatitis C virus
6	946	92.7	594	21	AAA73335	Hepatitis C virus
7	939	92.1	588	21	AAA73329	Hepatitis C virus
8	912	89.4	588	21	AAA73328	Hepatitis C virus
9	892.5	87.5	12734	24	ABA95615	Chimeric BVDV/HCV
10	885.5	86.8	1998	20	AAH80355	HCV NS4A-NS3 compl
11	882.5	86.5	1998	20	AAH80359	HCV NS4A-NS3 compl
12	881.5	86.4	1998	20	AAH80354	HCV NS4A-NS3 compl
13	878.5	86.1	651	20	AAH80345	HCV NS4A-NS3 compl
14	878.5	86.1	1998	20	AAH80358	HCV NS4A-NS3 compl
15	877.5	86.0	1998	20	AAH80353	HCV NS4A-NS3 compl
16	875.5	85.8	651	20	AAH80349	HCV NS4A-NS3 compl
17	874.5	85.7	612	25	ABX15706	Anti-viral synthet
18	874.5	85.7	651	20	AAH80344	HCV NS4A-NS3 compl
19	874.5	85.7	1998	20	AAH80357	HCV NS4A-NS3 compl
20	873.5	85.6	1998	20	AAH80352	HCV NS4A-NS3 compl
21	873.5	85.6	2013	20	AAH80360	HCV NS4A-NS3 compl
22	871.5	85.4	651	20	AAH80348	HCV NS4A-NS3 compl
23	870.5	85.3	651	20	AAH80343	HCV NS4A-NS3 compl
24	870.5	85.3	1998	20	AAH80356	HCV NS4A-NS3 compl
25	870.5	85.3	2016	20	AAH80361	HCV NS4A-NS3 compl
26	870	85.3	648	20	AAH80365	HCV NS4A-NS3 compl
27	868	85.1	648	20	AAH80363	HCV NS4A-NS3 compl
28	867.5	85.0	650	20	AAH80347	HCV NS4A-NS3 compl
29	867.5	85.0	651	20	AAH80351	HCV NS4A-NS3 compl
30	866.5	85.0	651	20	AAH80342	HCV NS4A-NS3 compl
31	864	84.7	648	20	AAH80362	HCV NS4A-NS3 compl
32	863.5	84.7	650	20	AAH80346	HCV NS4A-NS3 compl
33	863.5	84.7	651	20	AAH80350	HCV NS4A-NS3 compl
34	861	84.4	8145	20	AAH23259	Plasmid PET-BS(+)/
35	859	84.2	1933	20	AAH23258	HCV NS3 DNA. Hepa
36	858.5	84.2	9646	19	AAV59361	Hepatitis C virus
37	858.5	84.2	9646	24	ABK87285	Hepatitis C virus
38	858.5	84.2	12980	19	AAV59364	Hepatitis C virus
39	858.5	84.2	12980	24	ABK87286	Hepatitis C virus
40	858.5	84.2	16622	21	AZ36212	Nucleotide sequenc
41	854.5	83.8	5300	10	AAH92097	Combined open read
42	854.5	83.8	5360	10	AAH90327	Hepatitis C virus
43	854.5	83.8	6905	10	AAH92103	Combined open read
44	854.5	83.8	7310	10	AAH92106	Combined open read
45	854.5	83.8	7310	10	AAH90336	Composite hepatiti

ALIGNMENTS

RESULT 1

AAA73333

ID AAA73333 standard; DNA; 594 BP.

XX AAA73333;

XX

DT 19-DEC-2000 (first entry)

XX

DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #6.

XX

KW Hepatitis; NS3 protease; viral replication; chronic liver disease;

KW liver failure; liver cancer; mutant; mutein; ds.

XX

OS Hepatitis C virus.

OS Synthetic.

XX

FH Key Location/Qualifiers

PD 13-JUL-2000.
 XX
 PF 06-JAN-2000; 2000WO-US00345.
 PR
 PR 08-JAN-1999; 99US-0115271.
 XX
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX
 PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
 XX WPI; 2000-465976/40.
 DR P-PSDB; AAB15221.
 XX
 XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 PT
 XX
 PS Claim 26; Fig 13; 66pp; English.
 XX
 XX The present sequence is the coding sequence for a mutated version of a
 CC fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
 CC protease enzymes. These proteins are both essential for the replication
 CC of the virus, acting to cleave its replicative proteins from the
 CC polyprotein produced from the HCV genome. Inhibitors of the two proteins
 CC should be effective as antiviral treatments of HCV infection. This is
 CC useful as HCV can lead to chronic liver disease such as cirrhosis, liver
 CC failure and liver cancer. The present invention concerns a number of NS3
 CC mutants and NS3-NS4A fusion proteins which can be used to identify
 CC inhibitors of this type, as well as enabling structural studies of the
 CC protease and protease-inhibitor complexes. The protein produced from this
 CC sequence contains the alpha-helix0-1 variant.
 XX
 XX Sequence 594 BP; 103 A; 186 C; 156 G; 149 T; 0 other;
 SQ

Alignment Scores:
 Pred. No.: 1,87e-82 Length: 594
 Score: 973.00 Matches: 188
 Percent Similarity: 96.45% Conservative: 2
 Best Local Similarity: 95.43% Mismatches: 7
 Query Match: 95.39% Indels: 0
 DB: 21 Gaps: 0

US-09-965-594-20 (1-197) x AAA73330 (1-594)

QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
 DB 1 ATGAAAAAAGATCCGTTGTTATCGTCGGCGGTATCAACCTGTCCGGTGACACCGCT 60
 QY 21 TyrAlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThrGlyArgAsp 40
 DB 61 TAGCGTCAACGACTCGAGGTGAGGAGGGTTCGCAAGAAACCTCCAGACCGGTGCTGAC 120
 QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
 DB 121 AAAAACCGAGTTCAGGTGAAGTTCAGATCGTTCCACCGCTGCTCAGACCTTCCTGGCT 180
 QY 61 ThrSerIleAsnGlyValLeuThrThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
 DB 181 ACCTGTCATCAACGCTGTTGCTGGACCGTTTACACCGGTGCTGTACCGGTACCATCGCT 240
 QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr 100
 DB 241 TCCCGGAAGAGTCCGGTTATCCAGATGTCACCAACGTTGACAAAGACCTGGTGTGGTGG 300
 QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
 DB 301 CCGGCTCCGAGGGTCCCGTTCCTGACCCCGGTGACCTCGCGTTCCTCCGACCTGTAC 360
 QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
 DB 361 CTGGTTACCGGTACGGTACGTTATCCCGGTTTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 420

QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
 DB 421 CTGCTGTCCCGGTCGGATCTCTACTGAAAGGTTCCTCCGGTGGTCCGCTGCTGTC 480
 QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180
 DB 481 CCGGCTGGTCACGCTGTGGTATCTCCGTCGTGCTGTGTTCACCGCTGTGCTGCTAAA 540
 QY 181 AlalaValAspPheIleProValGluSerLeuLeuGluThrThrMetArgSerPro 197
 DB 541 GCIGTTGACTTCATCCCGGTTGAATCCCTGGAAACCAACCATGGTTCCTCCCG 591

RESULT 6
 AAA73335
 ID AAA73335 standard; DNA: 594 BP.
 XX
 AC AAA73335;
 XX
 DT 19-DEC-2000 (first entry)
 XX
 DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #8.
 XX
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
 KW liver failure; liver cancer; mutant; mutain; ds.
 OS
 OS Hepatitis C virus.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT CDS 1..594
 FT /*tag= a
 FT /product= "NS4A-NS3 fusion protein #8"
 XX
 PN WO200040707-A1.
 XX
 PD 13-JUL-2000.
 XX
 PF 06-JAN-2000; 2000WO-US00345.
 XX
 PR 08-JAN-1999; 99US-0115271.
 XX
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX
 PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
 XX WPI; 2000-465976/40.
 DR P-PSDB; AAB15226.
 XX
 PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 XX
 PS Disclosure; Fig 18; 66pp; English.
 XX
 XX The present sequence is the coding sequence for a mutated version of a
 CC fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
 CC protease enzymes. These proteins are both essential for the replication
 CC of the virus, acting to cleave its replicative proteins from the
 CC polyprotein produced from the HCV genome. Inhibitors of the two proteins
 CC should be effective as antiviral treatments of HCV infection. This is
 CC useful as HCV can lead to chronic liver disease such as cirrhosis, liver
 CC failure and liver cancer. The present invention concerns a number of NS3
 CC mutants and NS3-NS4A fusion proteins which can be used to identify
 CC inhibitors of this type, as well as enabling structural studies of the
 CC protease and protease-inhibitor complexes. The protein produced from this
 CC sequence contains the alpha-helix0 wild-type sequence.
 XX
 SQ Sequence 594 BP; 98 A; 189 C; 153 G; 154 T; 0 other;
 Alignment Scores:
 Pred. No.: 6.41e-80 Length: 594
 Score: 946.00 Matches: 186

Percent Similarity: 94.42% Conservative: 0
Best Local Similarity: 94.42% Mismatches: 11
Query Match: 92.75% Indels: 0
DB: 21 Gaps: 0

US-09-965-594-20 (1-197) x AAA73329 (1-588)

```
QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
DB 1 ATGAAAAAAGAGTCCGTTGTTATCGTCGCCGATCAACCTGTCGGTGACACCGCT 60
QY 21 TyrAlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThrGlyArgasp 40
DB 61 TACGCTCAGCAGACTCGAGGTCGCTGGTGTGCATCATCACTCCCTGACCGGTGCTGAC 120
QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
DB 121 AAAAACCCAGGTGAAGGTGAAGTTCAGATGTTTCCACCGCTGCTCAGACCTTCCTGGCT 180
QY 61 ThrSerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
DB 181 ACCTGCATCAACGGGTGTTGCTGGACCGTTTACACAGGTGCTGTACCGGTACCATCGCT 240
QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr 100
DB 241 TCCCGAAGGTCCCGTTATCCAGATGTACACCAACGTTGACAAGACCTGGTGGTGGTGG 300
QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
DB 301 CGCGCTCCGAGGTTCCCGTTCCTGACCCGCTGACCTCGCTGCTCCGACCTGTAC 360
QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
DB 361 CTGTTATCCCGTCAAGCTGACGTTATCCCGGTTCGTCGTGCTGCTCCCGGTGCTCC 420
QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyProLeuLeuLys 160
DB 421 CTGCTGTCCTCCCGCTCCGATCTCTACCTGAAAGGTTCCTCCGCTGCTGCTGCTGCT 480
QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaValSerThrArgGlyValAlaLys 180
DB 481 CGCGCTGCTCAGCTGTTGTTGTTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
DB 541 GCTGTTGATCTTATCCCGGTTGAATCTCCCTGGAACCAACCATGCGTTCCCGG 591
```

RESULT 7

AAA73329

ID AAA73329 standard; DNA; 588 BP.

XX AC AAA73329;

XX AC (first entry)

DT 19-DEC-2000

XX Hepatitis C virus NS4A-NS3 fusion protease coding sequence #2.

DE Hepatitis; NS3 protease; viral replication; chronic liver disease;

XX liver failure; liver cancer; mutant; mutin; ds.

XX Hepatitis C virus.

OS Synthetic.

XX Key Location/Qualifiers

PH 1..588

FT /tag= a

FT /product= "NS4A-NS3 fusion protein #2"

XX WO200040707-A1.

XX 13-JUL-2000.

XX 06-JAN-2000; 2000WO-US00345.

XX 08-JAN-1999; 99US-0115271.
XX (BRIM) BRISTOL-MYERS SQUIBB CO.
XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX WPI: 2000-465976/40.
XX P-PSDB; AAB15220.
XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
XX substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
XX amino acid, useful for screening inhibitors that may treat hepatitis C
XX
XX Claim 26; Fig 12; 66pp; English.
XX The present sequence is the coding sequence for a mutated version of a
XX fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
XX protease enzymes. These proteins are both essential for the replication
XX of the virus, acting to cleave its replicative proteins from the
XX polyprotein produced from the HCV genome. Inhibitors of the two proteins
XX should be effective as antiviral treatments of HCV infection. This is
XX useful as HCV can lead to chronic liver disease such as cirrhosis, liver
XX failure and liver cancer. The present invention concerns a number of NS3
XX mutants and NS3-NS4A fusion proteins which can be used to identify
XX inhibitors of this type, as well as enabling structural studies of the
XX protease and protease-inhibitor complexes. The protein produced from this
XX sequence contains the alpha-helix0-1 variant.
XX Sequence 588 BP; 103 A; 180 C; 156 G; 149 T; 0 other;

Alignment Scores:

Pred. No.: 2,88e-79 Length: 588
Score: 939.00 Matches: 184
Percent Similarity: 94.92% Conservative: 3
Best Local Similarity: 93.40% Mismatches: 8
Query Match: 92.06% Indels: 2
DB: 21 Gaps: 1

US-09-965-594-20 (1-197) x AAA73329 (1-588)

```
QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
DB 1 ATGAAAAAAGAGTCCGTTGTTATCGTCGCCGATCAACCTGTCGGTGACACCGCT 54
QY 21 TyrAlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThrGlyArgasp 40
DB 55 TACGCTCAGCAGACTCGAGTGGAGGTTGCCAAGAAACCTCCAGACCGGTGCTGAC 114
QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
DB 115 AAAAACCCAGGTGAAGGTGAAGTTCAGATGTTTCCACCGCTGCTCAGACCTTCCTGGCT 174
QY 61 ThrSerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
DB 175 ACCTGCATCAACGGGTGTTGCTGGACCGTTTACACGGGTGCTGTTACCGTACCATCGCT 234
QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr 100
DB 235 TCCCGAAGGTCCCGTTATCCAGATGTACACCAACGTTGACAAGACCTGTTGTTGGTGG 294
QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
DB 295 CGCGCTCCGAGGTTCCCGTTCCTGACCCCGTGCACCTGCGGTTCTCCGACCTGTAC 354
QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
DB 355 CTGCTTACCGCTCAGCTACGCTTATCCCGGTTCGTCGCTGCTGCTGCTGCTGCTGCT 414
QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyProLeuLeuLys 160
DB 415 CTGCTGTCCTCCCGCTCCGATCTCTACCTGAAAGGTTCCTCCCGTGGTCCGCTGCTGCTG 474
```

QY 161 ProAlaGlyHisValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180
 DB 475 CCGCGTGGTCACCGCTGTGTGTAATCTCCGCGCTGCTGTTGCACCGGTGGTGGCTAAA 534
 QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
 DB 535 GCTGTTGACTTCATCCCGGTGTAATCCCTGGAACCAACCATGCGTTCCCGG 585

RESULT 8
 AAA73328
 ID AAA73328 standard; DNA: 588 BP.
 XX
 AC AAA73328;
 XX
 DT 19-DEC-2000 (first entry)
 XX
 DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #1.
 XX
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
 KW liver failure; liver cancer; ds.
 XX
 OS Hepatitis C virus.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..588 /*tag= a
 FT /*product= "NS3-NS4A fusion protein"
 XX
 PN W0200040707-Al.
 XX
 PD 13-JUL-2000.
 XX
 PF 06-JAN-2000; 2000WO-US00345.
 XX
 PR 08-JAN-1999; 99US-0115271.
 XX
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX
 PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
 XX
 DR WPI: 2000-465976/40.
 DR P-PSDB: AAB15212.
 XX
 PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 PT
 XX
 PS Disclosure; Fig 10; 66pp; English.
 XX
 CC The present sequence is the coding sequence for a fusion protein created
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
 CC proteins are both essential for the replication of the virus, acting to
 CC cleave its replicative proteins from the polypeptide produced from the
 CC HCV genome. Inhibitors of the two proteins should be effective as
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A
 CC fusion proteins which can be used to identify inhibitors of this type, as
 CC well as enabling structural studies of the protease and
 CC protease-inhibitor complexes.
 XX
 SQ Sequence 588 BP; 97 A; 183 C; 153 G; 155 T; 0 other;

Alignment Scores:
 Pred. No.: 9.89e-77 Length: 588
 Score: 912.00 Matches: 182
 Percent Similarity: 92.89% Conservative: 1
 Best Local Similarity: 92.39% Mismatches: 12
 Query Match: 89.41% Indels: 2
 DB: 21 Gaps: 1

US-09-965-594-20 (1-197) x AAA73328 (1-588)
 QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
 DB 1 ATGAAAAAAGAGGTTCGGTGTATATCGTCGGCCGTATAGTACTGAACGGT-----GCT 54
 QY 21 TyrAlaGlnGlnThrArgGlyGluGlnLysThrSerHisThrGlyArgAsp 40
 DB 55 TACGCTCAGCAGACTCGAGGCTCGTGGTTCATCATCCTCCCTGACCGTGGTGGTAC 114
 QY 41 LysAsnGlnValGluGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
 DB 115 AAAAACCAGGTGAAGGTGAAGTTCAGATCGTTCCACCGCTGCTCAGACCTCTCTGGCT 174
 QY 61 ThrSerIleAspGlyValLeuThrThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
 DB 175 ACTGTCATCAACGGTGTTCGTGGACCGTTTACCACGGTGTGGTACCGGTACCATCGCT 234
 QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr 100
 DB 235 TCCCGCAAGGTCCGGTTATCCAGATGTACACCAACGTTGACAAAGACCTGGTGGTGG 294
 QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
 DB 295 CCGGCTCCGACGGTTCGGTTCCTGACCCGCTGACCTGCGGTTCCTCCGACCTGTAC 354
 QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
 DB 355 CTGGTTACCCGTCACGCTGACGTTATCCGGTTCGTCGTCGGTGGTACCTCCGTTGCC 414
 QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
 DB 415 CTGCTGTCCCGCTCCGATCTCTACCTGAAAGGTTCCTCCGGTGGTCCGCTGCTGTC 474
 QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180
 DB 475 CCGGCTGGTCAACGCTGTTGGTATCTCCGTCGTCGTCGGTGGTGGTGGTGGTGGT 534
 QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
 DB 535 GCTGTTGACTTCATCCCGGTGTAATCCCTGGAACCAACCATGCGTTCCCGG 585

RESULT 9

ABA95615

ID ABA95615 standard; DNA: 12734 BP.

XX

AC ABA95615;

XX

DT 21-MAR-2002 (first entry)

XX

DE Chimeric BVDV/HCV NS3-wt sequence.

XX

KW Pestivirus; Npro; protease; NS3; screening; ds.

XX

OS Chimeric - Bovine viral diarrhea virus.

XX

OS Chimeric - Hepatitis C virus.

XX

PN US6326137-B1.

XX

PD 04-DEC-2001.

XX

PF 25-JUN-1999; 99US-0344456.

XX

PR 25-JUN-1999; 99US-0344456.

XX

PA (SCHE) SCHERING CORP.

XX

PI Hong Z, Lai VCH, Lau JYN;

XX

DR WPI: 2002-121103/16.

XX

PT Nucleic acid construct encoding chimeric Hepatitis C Virus (HCV)

QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
 DB 304 CAAAGGGCCATCACCCAGATGTACACTAATGTGGACAGACCTCTCGGCTGGCAG 363
 QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
 DB 364 GCGCCCCCGGGGGGCTCTCTTGACACCATGACCTGTGGCAGCTCAGACCTTTACTTG 423
 QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
 DB 424 GTCACGAGACATGCTGACGTATTCGGTCCGGCGGGGGGACAGTAGGGGAGCCCTG 483
 QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
 DB 484 CTCTCCCCAGGCTGCTCTCTACTTTGAAGGGCTCTTCGGGTGGTCCACTGCTCTGCCCT 543
 QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
 DB 544 TCGGGGACGCTGTGGGCATCTTCGGGGCTGCGGTATGCACCGGGGGGTTCGGAAGCG 603
 QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
 DB 604 GTGGACTTTGTGCGGTAGACTCCATGGAACACTATGCGGTCTCCG 651

RESULT 11

AAx80359
 ID AAX80359 standard; cDNA; 1998 BP.

XX AC AAX80359;

XX DT 07-SEP-1999 (first entry)

XX DE HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:109.

XX KW HCV; hepatitis C virus; single chain recombinant complex; linker;

XX KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;

XX KW hydrophobic domain; covalent complex; detection; inhibitor; ss.

XX OS Hepatitis C virus.

XX OS Synthetic.

XX PN W09928482-A2.

XX PD 10-JUN-1999.

XX PF 24-NOV-1998; 98WO-US24528.

XX PR 28-JUL-1998; 98US-0094331.

XX PR 28-NOV-1997; 97US-0067315.

XX PA (SCHE) SCHERING CORP.

XX PI Malcolm BA, Taremi SS, Weber PC, Yao N;

XX DR WPI; 1999-385385/32.

XX PT New hepatitis C virus covalent complexes

XX PS Disclosure; Page 179-182; 21pp; English.

XX CC The present invention describes a covalent hepatitis C virus (HCV)

XX CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV

XX CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the

XX CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker

XX CC to the amino terminus of the HCV NS3 protease domain. The present

XX CC sequence encodes an example of the above complex. The covalent

XX CC NS4A-NS3 complexes are useful for structural determination and

XX CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.

XX CC They can also be used for detecting inhibitors of the protease activity,

XX CC the helicase activity and the ATPase activity of NS3. The covalent

XX CC NS4A-NS3 complexes are more soluble, stable and active than the non-

XX CC covalent protease-peptide complexes previously available.

SQ Sequence 1998 BP: 411 A; 595 C; 569 G; 423 T; 0 other;

Alignment Scores:

Pred. No.: 2,69e-73 Length: 1998
 Score: 882.50 Matches: 167
 Percent Similarity: 93.37% Conservative: 16
 Best Local Similarity: 85.20% Mismatches: 10
 Query Match: 86.52% Indels: 3
 DB: 20 Gaps: 1

OS-09-965-594-20 (1-197) x AAX80359 (1-1998)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21

DB 64 GGTCTGTGTTATTGTTGTAGAAATTATTTATCTGTAGTGTAGTATCATCGGCTAC 123

QY 22 AlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThrGlyArgAspLys 41

DB 124 TCCCAACAGACGCGGGGCTACTTTGGTTGCAAGAGACTAGCCTTACAGCGCGGACAG 183

QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61

DB 184 AACCAAGTTCGAGGAGAGGTTTCAGTGGTTCCACCGCACACAAATCCTTCCTGGCGACC 243

QY 62 SerIleAsnGlyValLeuTyrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81

DB 244 TCGGTCAACGCGCTGTGTGGACGCTTTACCATGCTGTGCTCAAGACCTTAGCGCGC 303

QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101

DB 304 CCAAGGGGCGCAATCACCAGATGTACACTAATGTGGACAGGACCTCGTGGCTGCAG 363

QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121

DB 364 GCGCCCCCGGGGCGCTCTCTGACACCATGCACCTGTGGCAGCTCAGACCTTTACTTG 423

QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141

DB 424 GTCACGACATGCTGACGCTATTCGGTGGCGGGGGGCGACAGTAGGGGAGCGCTG 483

QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161

DB 484 CTCTCCCCAGGCTGCTCTCTACTTTGAAGGGCTCTGCTGGTGTCCACTGCTCTGCCCT 543

QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181

DB 544 TCGGGGACGCTGTGGGCATCTTCGGGCTGCGGTATGCACCGGGGGGTTCGGAAGCG 603

QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197

DB 604 GTGGACTTTGTGCGGTAGACTCCATGGAACACTACTATGCGGTCTCCG 651

RESULT 12

AAx80354

ID AAX80354 standard; cDNA; 1998 BP.

XX AC AAX80354;

XX DT 07-SEP-1999 (first entry)

XX DE HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:104.

XX KW HCV; hepatitis C virus; single chain recombinant complex; linker;

XX KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;

XX KW hydrophobic domain; covalent complex; detection; inhibitor; ss.

XX OS Hepatitis C virus.

XX OS Synthetic.

XX PN W09928482-A2.

XX PD 10-JUN-1999.

```

PF 24-NOV-1998; 98WO-US24528.
XX
PR 28-JUL-1998; 98US-0094331.
PR 28-NOV-1997; 97US-0067315.
XX
PA (SCHE ) SCHERING CORP.
XX
PI Malcolm BA, Taremi SS, Weber PC, Yao N;
XX
DR WPI; 1999-385385/32.
XX
XX New hepatitis C virus covalent complexes
XX
PS Disclosure; Page 163-166; 21lpp; English.
XX
CC The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence encodes an example of the above complex. The covalent
CC NS4A-NS3 complexes are useful for structural determination and
CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.
CC They can also be used for detecting inhibitors of the protease activity,
CC the helicase activity and the ATPase activity of NS3. The covalent
CC NS4A-NS3 complexes are more soluble, stable and active than the non-
CC covalent protease-peptide complexes previously available.
XX
SQ Sequence 1998 BP; 410 A; 596 C; 568 G; 424 T; 0 other;

Alignment Scores:
Pred. No.: 3.34e-73 Length: 1998
Score: 881.50 Matches: 168
Percent Similarity: 92.86% Conservative: 14
Best Local Similarity: 85.71% Mismatches: 11
Query Match: 86.42% Indels: 3
DB: 20 Gaps: 1

US-09-965-594-20 (1-197) x AAX80354 (1-1998)
Oy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
Db 64 GGTCTCTGTTATGTTGGTAGAATATTTATCTGGTAGTGGTAGTATCATCGGCTAC 123
Oy 22 AlaGlnGlnThrArgGlyGluGlnGlnGlyCysGlnLysThrSerHisThrGlyArgAspLys 41
Db 124 TCCCAACAGACAGCGGGCCCTACTTGGTGCATCAAGACTAGCCTTACAGCGCGGCAAG 183
Oy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
Db 184 AACAGGTCAGGGAGAGGTTTCAGTGGTTTCACCGCAACACAACTCTCTGGCGACC 243
Oy 62 SerIleAsnGlyValLeuThrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
Db 244 TGCCTCAACGCGGTGTGTGGACCGTTTACCATGGTGTGGCTGCTCAAGACCTTAGCGGC 303
Oy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101
Db 304 CCAAGGGGCGCAATCACCCAGATGTACACTATGTGGACCGACGACCTCTGGCTGGCAG 363
Oy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuLysLeu 121
Db 364 CGCGCCCGCGGGCGCGCTCTCTTGACACCATGTCACCTGTGGCAGCTCAGACCTTTACTTG 423
Oy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
Db 424 CTCACGACATGTGTACGTCAATTCGGGTGCGCGGGCGGCGACAGTAGGGGGAGCCCTG 483
Oy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
Db 484 CTCCTCCCGCCAGCCCTGTCTCTACTTCAAGGGGCTCTTCGGGTGGTCCACTGCTCTGCCCT 543
Oy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181

```

```

Db 544 TCGGGCAGCGTGTGGGCATCTTCGGGCTGCGGTATGCACCGGGGGTTCGAGGCG 603
Oy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
Db 604 GTGGACTTGTGCGCGTAGAGTCCATGTGAAGACTACTATGCGGTCTCCG 651

RESULT 13
AAX80345
ID AAX80345 standard; cDNA; 651 BP.
XX
AC AAX80345;
XX
DT 07-SEP-1999 (first entry)
XX
DE HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:95.
XX
CC HCV; hepatitis C virus; single chain recombinant complex; linker;
CC NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
CC hydrophobic domain; covalent complex; detection; inhibitor; ss.
XX
OS Hepatitis C virus.
XX
XX Synthetic.
XX
FN W09928482-A2.
XX
PD 10-JUN-1999.
XX
PF 24-NOV-1998; 98WO-US24528.
XX
PR 28-JUL-1998; 98US-0094331.
PR 28-NOV-1997; 97US-0067315.
XX
PA (SCHE ) SCHERING CORP.
XX
PI Malcolm BA, Taremi SS, Weber PC, Yao N;
XX
DR WPI; 1999-385385/32.
XX
XX New hepatitis C virus covalent complexes
XX
PS Disclosure; Page 147-148; 21lpp; English.
XX
CC The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence encodes an example of the above complex. The covalent
CC NS4A-NS3 complexes are useful for structural determination and
CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.
CC They can also be used for detecting inhibitors of the protease activity,
CC the helicase activity and the ATPase activity of NS3. The covalent
CC NS4A-NS3 complexes are more soluble, stable and active than the non-
CC covalent protease-peptide complexes previously available.
XX
SQ Sequence 651 BP; 120 A; 187 C; 200 G; 144 T; 0 other;

Alignment Scores:
Pred. No.: 1.58e-73 Length: 651
Score: 878.50 Matches: 167
Percent Similarity: 93.33% Conservative: 15
Best Local Similarity: 85.64% Mismatches: 10
Query Match: 86.13% Indels: 3
DB: 20 Gaps: 1

US-09-965-594-20 (1-197) x AAX80345 (1-651)
Oy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
Db 64 GGTCTCTGTTATGTTGGTAGAATATTTATCTGGTAGTGGTAGTATCATCGGCTAC 123
Oy 22 AlaGlnGlnThrArgGlyGluGlnGlnGlyCysGlnLysThrSerHisThrGlyArgAspLys 41

```

```

Db 124 TCCCAACAGCGGGCCCTACTGGTTCCAAAGACACTAGCCTTACAGCGCGGACACAG 183
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
Db 184 ACCAGGTGAGGAGAGGTTGAGGTGGTTTCCACCGCAACAAATCTTCTTGGCGACC 243
QY 62 SerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
Db 244 TCGGTCAACGCGGTGTGTGGACCGTTTACCAATGTGTGCTGGCTCAAGACCTTAGCCGCG 303
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
Db 304 CCAAGGGGCCAATCAACCCAGATGTACATAATGTGGACCAAGGACCTCGCTGGCTGGCAG 363
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
Db 364 GCGCCCGCGGGCGGTCTTGTACACCATGTCACCTGTGGCAGTCACACCTTTACTTG 423
QY 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141
Db 424 GTCACGAGACATGCTCACGTCATTCGGGTGCGCGCGCGGCGGACAGTAGGGGAGCCCTG 483
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
Db 484 CTCCTCCCGCAGGCGCTGTCTCTACTTTGAAGGGCTCTTCGGGTGGTCCACTGCTCGCCT 543
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
Db 544 TCGGGCAGCGTGTGGCATCTTCGGGCTGCGGTATGCACCGGGGGTTCGGAAGCG 603
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196
Db 604 GTGACCTTTGTGCGCGTAGAGTCCATGGAAACTACTATGCGGTCT 648

```

RESULT 14

```

AA80358
ID AAX80358 standard; cDNA; 1998 BP.
XX
AC AAX80358;
XX
DT 07-SEP-1999 (first entry)
XX
DE HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:108.
XX
KW HCV; hepatitis C virus; single chain recombinant complex; linker;
KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
KW hydrophobic domain; covalent complex; detection; inhibitor; ss.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
XX WO9928482-A2.
XX
XX 10-JUN-1999.
XX
XX 24-NOV-1998; 98WO-US24528.
XX
XX 28-JUL-1998; 98US-0094331.
XX
XX 28-NOV-1997; 97US-0067315.
XX
XX (SCHE ) SCHERING CORP.
XX
XX Malcolm BA, Taremi SS, Weber PC, Yao N;
XX WPI; 1999-385385/32.
XX
XX New hepatitis C virus covalent complexes
XX
XX Disclosure; Page 176-179; 21pp; English.
XX
XX The present invention describes a covalent hepatitis C virus (HCV)
XX NS4A-NS3 complex comprising a central hydrophobic domain of native HCV

```

```

CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence encodes an example of the above complex. The covalent
CC NS4A-NS3 complexes are useful for structural determination and
CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.
CC They can also be used for detecting inhibitors of the protease activity,
CC the helicase activity and the ATPase activity of NS3. The covalent
CC NS4A-NS3 complexes are more soluble, stable and active than the non-
CC covalent protease-peptide complexes previously available.
XX
SQ Sequence 1998 BP: 410 A; 596 C; 568 G; 424 T; 0 other;

```

Alignment Scores:

```

Pred. No.: 6,39e-73 Length: 1998
Score: 878.50 Matches: 167
Percent Similarity: 92.86% Conservative: 15
Best Local Similarity: 85.20% Mismatches: 11
Query Match: 86.13% Indels: 3
DB: 20 Gaps: 1

```

US-09-965-594-20 (1-197) x AAX80358 (1-1998)

```

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
Db 64 GGTTCGTGTGTATTGTTGGTAGAATTATTTTATCTGGTAGTGTATATCAGCGCCTAC 123
QY 22 AlaGlnGlnThrArgGlyGluGlnCysGlnLysThrSerHisThrGlyArgAspLys 41
Db 124 TCCCAACAGACGCGGGGCTACTTGTGTGCATCAAGACTAGCCTTACAGCGCGGACAG 183
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
Db 184 AACCAAGTCGAGGAGAGGTTTCAGGTGTTTCCACCGCAACAAATCTTCTCGCGACC 243
QY 62 SerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
Db 244 TCGGTCAACGCGGTGTGTGTGGACCGTTTACCATGTGTGGCTCAAGACCTTAGCCGCG 303
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
Db 304 CCAAGGGGCCAATCAACCCAGATGTACATAATGTGGACCAAGGACCTGCTGGCTGGCAG 363
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
Db 364 GCGCCCGCGGGCGGTCTTGTACACCATGTCACCTGTGGCAGTCACACCTTTACTTG 423
QY 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141
Db 424 GTCACGAGACATGCTGACGTATTCGGGTGCGCGCGGCGGACAGTAGGGGAGCGCTG 483
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
Db 484 CTCCTCCCGCAGGCGCTGTCTCTACTTTGAAGGGCTCTGCTGGTGGTCCACTGCTCGCCT 543
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
Db 544 TCGGGCAGCGTGTGGCATCTTCCGGCTGCGGTATGCACCGCGGGGTTCGGAAGCG 603
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
Db 604 GTGACCTTTGTGCGCGTAGAGTCCATGGAAACTACTATGCGGTCTCCG 651

```

RESULT 15

```

AAX80353
ID AAX80353 standard; cDNA; 1998 BP.
XX
AC AAX80353;
XX
DT 07-SEP-1999 (first entry)
XX
XX HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:103.
XX

```

KW HCV; hepatitis C virus; single chain recombinant complex; linker;
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
 KW hydrophobic domain; covalent complex; detection; inhibitor; ss.
 XX Hepatitis C virus.
 OS Synthetic.

XX WO928482-A2.
 XX 10-JUN-1999.

XX PF 24-NOV-1998; 98WO-US24528.
 XX PR 28-JUL-1998; 98US-0094331.
 XX PR 28-NOV-1997; 97US-0067315.

XX PA (SCHE) SCHERING CORP.
 XX PI Malcolm BA, Taremi SS, Weber PC, Yao N;
 XX WI: 1999-385385/32.

XX PT New hepatitis C virus covalent complexes
 XX PS Disclosure: Page 160-162; 21pp; English.

XX The present invention describes a covalent hepatitis C virus (HCV)
 CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
 CC to the amino terminus of the HCV NS3 protease domain. The present
 CC sequence encodes an example of the above complex. The covalent
 CC NS4A-NS3 complexes are useful for structural determination and
 CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.
 CC They can also be used for detecting inhibitors of the protease activity,
 CC the helicase activity and the ATPase activity of NS3. The covalent
 CC NS4A-NS3 complexes are more soluble, stable and active than the non-
 CC covalent protease-peptide complexes previously available.

XX SQ Sequence 1998 BP; 410 A; 596 C; 568 G; 424 T; 0 other;

Alignment Scores:
 Pred. No.: 7,93e-73 Length: 1998
 Score: 877.50 Matches: 167
 Percent Similarity: 92.86% Conservative: 15
 Best Local Similarity: 85.20% Mismatches: 11
 Query Match: 86.03% Indels: 3
 DB: 20 Gaps: 1

US-09-965-594-20 (1-197) x AAX80353 (1-1998)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
 DB 64 GGTCTGTGTATTGTTGGTAGAATATTTATCTGGTAGTGGTAGTATCATCGGCTAC 123
 QY 22 AlaGlnGlnThrArgGlnGlnGlnCysGlnLysThrSerHisThrGlyArgAspLys 41
 DB 124 TCCCAACAGACCGCGGGCTACTTGGTGGCAAGATCATTACCTTACAGCGCGGACAG 183
 QY 42 AsnGlnValGlnGluGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
 DB 184 AACCAAGTTCGAGGGAGAGTTTCAGGTGGTTTCCACCGCAACAACATCTCTCTGGCGACC 243
 QY 62 SerIleAsnGlyValLeuThrValThrHisGlyAlaGlyThrArgThrIleAlaSer 81
 DB 244 TCGGTCAACGGCGTGTGTGGACCGTTTACCATTGGTGGCTCAAGACCTTAGCCGGC 303
 QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
 DB 304 CCNAGGGGCCAATCACCAGATGTACTAATGTGGACAGGACCTCGTGGCTGGCAG 363
 QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121

DB 364 GCGCCCCCGGGCGGCTTCCTTGACACCATGCACCTGTGGCAGCTCAGACCTTTACTTG 423
 QY 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141
 DB 424 GTCACGAGACATGCTGACGTCAITCCGGTGCCTCCGCGGGGCGACAGTAGGGGAGCCTG 483
 QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
 DB 484 CTCTCCCCCAGGCTGTCTCTCTACTTGAAGGGCTCTTCGGGTGGTCCACTGCTCTGCCCT 543
 QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
 DB 544 TCGGGGCACGCTGTGGGCATCTTCCGGGTGCCGTATGCACCCGGGGGTTGCCAAGCGC 603
 QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
 DB 604 GTGGACTTTGTGCCCGTAGAGTCCATGGNAACTACTATGCGGTCTCCG 651

Search completed: August 30, 2003, 19:48:15
 Job time : 188.939 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 30, 2003, 19:20:43 ; Search time 1910.31 Seconds
(without alignments)
2506.388 Million cell updates/sec

Title: US-09-965-594-20
Perfect score: 1020
Sequence: 1 MKKKGSVVIGRINLSGDTA.....VAKAVDFIPVESLETTMRSP 197

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-Q/cgn2.1/USPTO_spool/US09965594/runat_29082003_151919_28322/app_query.fasta_1.2872
-DB-EST -OFMT-fastap -SUFFIX-rst -MINMATCH-0.1 -LOOPEL-0 -LOOPEXT-0
-UNITS-bits -START-1 -END-1 -MATRIX-blosum62 -TRANS-human40 cdi -LIST-45
-DOCALIGN-200 -THR_SCORE-pct -THR_MAX-100 -THR_MIN-0 -ALIGN-15 -MODE-LOCAL
-OUTFMT-pco -NORM-ext -HEAPSIZE-500 -MINLEN-0 -MAXLEN-2000000000
-USER-US09965594@cgn_1.1.12630.#runat_29082003_151919_28322 -NCPU-3
-NO_MAP -LARGEQUERY -NEG_SCORES-0 -WAIT -DSPBLOCK-100 -LONGLOG
-DEV_TIMEOUT-120 -WARN_TIMEOUT-30 -THREADS-1 -XGAPOP-10 -XGAPEXT-0.5 -FCGAPOP-6
-FCGAPEXT-7 -YGAPOP-10 -YGAPEXT-0.5 -DELOP-6 -DELEXT-7

Database : EST:.*
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_hic.*
9: gb_est1.*
10: gb_est2.*
11: gb_hic.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: em_gss_hum.*
18: em_gss_inv.*
19: em_gss_pln.*
20: em_gss_vrt.*
21: em_gss_fun.*
22: em_gss_nam.*
23: em_gss_mus.*
24: em_gss_pro.*
25: em_gss_rod.*
26: em_gss_phg.*
27: em_gss_vrl.*
28: gb_gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	106.5	10.4	1031	14	CB950999
2	106	10.4	1146	12	BM915803
3	102.5	10.0	1403	13	BQ926101
4	101	9.9	1199	13	BQ892487
5	97.5	9.6	629	10	BG089727
6	97	9.5	644	29	BX238988
7	95.5	9.4	701	10	BF863244
8	95.5	9.4	984	10	BF304699
9	94	9.2	560	28	A0538021
10	94	9.2	935	10	B868757
11	94	9.2	1062	29	CHS060HN
12	93.5	9.2	772	29	CC406704
13	93.5	9.2	789	29	CC406705
14	93	9.1	528	12	BM402566
15	93	9.1	701	29	B2342381
16	93	9.1	1213	13	B0541777
17	92.5	9.1	938	13	B0894657
18	91.5	9.0	502	9	AA036834
19	91.5	9.0	528	28	AQ620249
20	91.5	9.0	812	13	B0299264
21	91.5	9.0	817	13	B0240438
22	91.5	9.0	878	13	B0365755
23	91	8.9	574	29	CG380642
24	91	8.9	579	29	CG380645
25	91	8.9	580	14	CAT28398
26	91	8.9	753	13	B0402910
27	91	8.9	866	13	B0219343
28	91	8.9	905	13	B0542842
29	91	8.9	906	13	BX434207
30	91	8.9	936	29	CG373208
31	90.5	8.9	622	9	AV835401
32	90.5	8.9	814	11	CHS09179
33	90.5	8.9	824	13	B0396324
34	90.5	8.9	958	10	BG420860
35	90.5	8.9	1035	10	BE888775
36	90.5	8.9	1141	11	AK080545
37	90.5	8.9	1440	12	BM467279
38	90.5	8.9	1733	12	BM553374
39	90.5	8.9	3215	11	AK051518
40	90	8.8	500	12	BM708007
41	90	8.8	569	12	BM825317
42	90	8.8	617	10	BE05938
43	90	8.8	631	10	AK961059
44	90	8.8	658	12	BM830847
45	90	8.8	757	12	B1258851

ALIGNMENTS

RESULT 1
CB950999
LOCUS CB950999
DEFINITION AGENCOURT_13445496 NIH_MGC_177 Mus musculus CDNA clone
IMAGE:30316162 5', mRNA sequence.
ACCESSION CB950999
VERSION CB950999.1 GI:30205777
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1031)


```

QY 61 rSerIleAsnGlyValLeuTrpThrValTyrHis-----GlyAlaGlyThr-- 76
Db 1002 CCGCGCTGGTGGATGTGTGG-----TATCACTTCCCGCGCGGGGAGGAGTACGTG 949
QY 77 -----ArgThrIleAlaSerPro-----LysGlyProValThrGlnMetTy 90
Db 948 ACCGAGGGGCGCGCGTCCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 903
QY 90 rThrAsnValAspLysAspLeuValGlyTrpGlnAlaProGln-----GlySerAr 107
Db 902 -----CAGATGTGCGGTGGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 859
QY 107 gSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeuValThrArgHisAlaAs 127
Db 858 ACTTCTGTCTGCTGTTCTGTGG-----TATCACTTCCCGCGCGGGGAGGAGTACGTG 949
QY 127 pValIleProValArgArgGlyAspSerArgGlySerLeuLeuSerProArgProI 147
Db 833 -----CGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 787
QY 147 eSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysProAlaGlyHisAlaValG 167
Db 786 CCGGTATCTACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 742
QY 167 yIlePheArgAlaAlaValSerThrArgGlyValAlaLysAlaValAspPhe---IlePr 186
Db 741 GCGCTTCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 682
QY 186 oVal 187
Db 681 TTTG 678

RESULT 3
BQ926101/c
LOCUS BQ926101.1 1403 bp mRNA linear EST 20-AUG-2002
DEFINITION AGENCOURT_8752655 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6335718
5', mRNA sequence.
ACCESSION BQ926101
VERSION BQ926101.1 GI:22341132
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1403)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,
Ph.D.
cDNA Library Preparation: ResGen, Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L14M13798 row: j column: 07
High quality sequence stop: 101.
Location/Qualifiers
1. .1403
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6335718"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_130"
/note="Organ: oCocysts; Vector: pCMV-SPORT6.1.cdb;
Site_1: EcoRV; Site_2: NotI; Cloned unidirectionally.
Primer: Oligo dt. Average insert size 1.95 kb.
Constructed by ResGen, Invitrogen Corp. Note: this is a

```

```

NIH_MGC Library..
BASE COUNT 297 a 521 c 237 g 345 t 3 others
ORIGIN
Alignment Scores:
Pred. No.: 8.34 Length: 1403
Score: 102.50 Matches: 58
Percent Similarity: 36.00% Conservative: 14
Best Local Similarity: 29.00% Mismatches: 66
Query Match: 10.05% Indels: 62
DB: 13 Gaps: 11
US-09-965-594-20 (1-197) x BQ926101 (1-1403)
QY 11 GlyArgIleAsnLeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGlnGly 30
Db 1378 GGGTGTGTTTCANCGGTACGACAGGTGCGCC--GCACACTCGACCGCTCGCGCAGACT 1322
QY 31 CysGlnLysThr-----SerHisThrGlyArgAspLysAsnGln-----Val 44
Db 1321 TGTGCGGGGCGCGTTCGCGCATACCGCGGTGCGAGTTCAGGCGCGCTGTATACA 1262
QY 45 GluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThrSerIleAsn 64
Db 1261 GAGGGGAAA-----CAG 1250
QY 65 GlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLys--G 84
Db 1249 GGGGTA---TGGTTATCACGGGCTGGGCGAGGTACT-----TCCCTTAAAGCG 1205
QY 84 lyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGlnAlaProG 104
Db 1204 CGCGCGTGGCGAGTATATATACCGCGGAGTGCAGAACCGACGGCGGTGGAACGTTGACC 1145
QY 104 lnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeuValThrA 124
Db 1144 AA---GAGAGCGACCTGACCGCCCTCCCTGTGGGCTGCGATATAACAATATGTCAG 1088
QY 124 rgHisAlaAspValIleProValArgArgGlyAsp-----136
Db 1087 GGCACGGTGTGTTACTACCGCGCAGACCGCTCCACACGGCGCTCTCTAACACACGC 1028
QY 137 -----SerArgGlySerLeuSerProArgProIle-SerTyrLeuLysGlySerSer 154
Db 1027 CCGCGCTCCCGCGCAAC-----AGGTAATAATCATATCGCGCGGGGATTTC 980
QY 155 Gly-----GlyProLeuLeuCys 160
Db 979 GCATTCGCGGGGAGAGCGCGGTGCGGGGCGCGCGCTCGCGCGCGCTGAGGCGC 920
QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyVal 178
Db 919 AGGAGAGGC-----GCGGTGTTTCGCGGGGTGAGGACGAAGCGCGCGGTG 875
RESULT 4
BQ92487
LOCUS BQ92487
DEFINITION AGENCOURT_8417538 Lupski_sympathetic_trunk Homo sapiens cDNA clone
IMAGE:6192708 5', mRNA sequence.
ACCESSION BQ92487
VERSION BQ92487.1 GI:22284501
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 1199)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. James R. Lupski

```

cdna Library Preparation: Life Technologies, Inc.
 cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MCC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM13595 row: c column: 13
 High quality sequence start: 57
 High quality sequence stop: 394.

FEATURES

source
 1. .1199
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6192708"
 /sex="male"
 /tissue_type="sympathetic trunk"
 /dev_stage="adult, 16 yr"
 /lab_host="DH10B"
 /clone_lib="Lupski_sympathetic_trunk"
 Note: Vector: pCMV-SPORT6 (Life Technologies); Site_1:
 Not1: Site_2: SalI: cdna made by oligo-dt priming.
 Directionally cloned using the following adaptors:
 5'-TCGACGACGCGGCGG-3' and
 5'-GACGAGTCTGATGCGGCGGCGG-3'. Size selected >
 1 kb for average insert length 1.9 kb. This is a primary
 library, non-amplified. Library constructed by Life
 Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
 College of Medicine); available through Life
 Technologies.
 BASE COUNT 255 a 362 c 343 g 211 t 28 others
 ORIGIN

Alignment Scores:
 Pred. No.: 9.59 Length: 1199
 Score: 101.00 Matches: 50
 Percent Similarity: 34.74% Conservative: 24
 Best Local Similarity: 23.47% Mismatches: 80
 Query Match: 9.90% Indels: 59
 DB: 13 Gaps: 9

US-09-965-594-20 (1-197) x BQ892487 (1-1199)

QY 16 SerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSer 35
 DB 337 GCAGGAGAGAACCTTACCCCAACAG-----AAGGCA 369
 QY 36 HistHrGlyArgAspLysAsnGlnValGluGluValGlnLleValSerThr----- 53
 DB 370 CATGGGGAAATCGCCGCTTCAGAGACGAGGTGCTTCATGTTTCTGAAACATAACCG 429
 QY 54 AlaThrGlnThrPheLeu-----AlaThrSerIleAsnGlyValLeuThr 69
 DB 430 CCAGCCACTGCTTCATGTAATGATACCCCTTCCACCACACACAGGGCAGCATGGGAT 489
 QY 70 ValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLysGlyProValThrGlnMet 89
 DB 490 CCATTTTAAAGGGTGCTCTGTTAATCATGCGCCGCGCCGCGCTGATCTCCA 549
 QY 90 TyrThrAsnValAspLysAspLeuValGlyTrpGlnAlaProGlnGlySerArgSerLeu 109
 DB 550 TTTACCATGTGACAGTGACTTT-----TGT 576
 QY 110 ThrProCysThr-----CysGlySerSerAspLeuTyr 120
 DB 577 GTCGCTGCACAGACACCCCATGACCGATGTTGGCTTATGTGGAACGGCGGCGGTC 636
 QY 121 -LeuValThr-----ArgHisAlaAspValIleProValArg----- 132
 DB 637 ATGGCCACTCCCTCTATTAACACACGCGAAGCTGTTCCATGGCGCGGCGGTGTGT 696
 QY 133 -----ArgArgGlyAspSerArgGlySerLeuLeu----- 142

Db 697 TTGCAGCGCAAGCGGGTGGGGCTAGGACTCGGGGGCGGATCTCTGAAACC 756
 QY 143 -SerProArgProIleSerTyrLeuLys-----GlySerSerGlyGlyPr 157
 Db 757 CCACCTCGGGCCACCGCATGCGCTTAAGCCTCCCTTTTACAAGCCACCGCGCGCCGCC 816
 QY 157 oLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgG 177
 Db 817 CTTACATCTCTACCTCGGCGCGGGGGGAGAGCTGGGGCGCATACGGCTCAGGG 876
 QY 177 yValAlaLysAlaValAspPheIleProValGluSer 189
 Db 877 CGTTTAAAGCCCGCGCTTCGCGCGCGGCGGAAGCA 913

RESULT 5

BQ89727/c
 LOCUS BQ89727.1 629 bp mRNA linear EST 26-JAN-2001
 DEFINITION mab90e06.x1 NCI_CGAP_Sp2 Mus musculus cdna clone IMAGE:3977578 3'
 similar to SW:GRAD_MOUSE P11033 GRANZYME D PRECURSOR ;, mRNA
 sequence.
 ACCESSION BQ89727
 VERSION BQ89727.1 GI:12572290
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 629)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLES National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-re@mail.nih.gov
 Tissue Procurement: David Segal Ph.D., Herbert Morse M.D.
 cdna Library Preparation: Life Technologies, Inc.
 cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov

MGI:1477610

Seq primer: -40UP from Gibco
 High quality sequence stop: 422.

FEATURES

source
 1. 629
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:3977578"
 /tissue_type="NK cells (flow-sorted)"
 /lab_host="DH10B (TI-resistant)"
 /clone_lib="NCI_CGAP_Sp2"
 /note="Organ: spleen; Vector: pCMV-SPORT6 (Life Technologies); mRNA made from flow-sorted NK cells, cdna made by oligo-dt priming. Directionally cloned. Average insert size 1.5 kb. Primary library, non-amplified. cdna Library Preparation: David B. Krizman, Ph.D."
 BASE COUNT 131 a 156 c 150 g 191 t 1 others
 ORIGIN

Alignment Scores:

Pred. No.: 9.21 Length: 629
 Score: 97.50 Matches: 48
 Percent Similarity: 39.23% Conservative: 23
 Best Local Similarity: 26.52% Mismatches: 59
 Query Match: 9.56% Indels: 51
 DB: 10 Gaps: 12

US-09-965-594-20 (1-197) x BQ89727 (1-629)

QY 36 HistHrGlyArgAspLysAsnGlnValGluGlyGluValGlnLleValSerThrAlaThr 55

```

Db 620 CATCGGGTAAG-----GAAGGAGACACAGATCATCCCTTGTGCA--- 579
QY 56 GlnThrPheLeuAlaThrSerIleAsnGlyValLeuThrValThrHisGly----- 73
Db 578 AAACATTTCCTCCAGATATAAATGCT-----ACTATCTCTTCAGGTGAGATC 528
QY 74 -----AlaGlyThrArgThrIleAlaSer 81
Db 527 ATGCTGTTAAAGCTGGAGAGTAAGCCAGAGAACTAAAGCTGTGAGACCCCTCAAGTTG 468
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrPln 101
Db 467 CCCAGATCCAAATGCCCGGCTGAAGCCAGGAATGTG---TGCAGTGTGCTGGCTGG--- 414
QY 102 AlaProGlnGlySerArgSerLeu-----ThrProCysThrCysGlySerSerAspLeu 119
Db 413 -----GGGTCAAGTCCCATCAATGACACATCAAGCATCTGCCGCTCCGAGAGGTT 363
QY 120 TyrLeuValThrArgHisAlaAspValIleProValArgArgArg----- 134
Db 362 CAACCTGGTTCATCCAGGAGGACGAGGAATGCAAAAAACGTTTCCGATACACTAGACAC 303
QY 135 -----GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyr 149
Db 302 ACAGAGATTGTGCTGGAGACTTGAAG---AAAAATNAAGACTCCT----- 261
QY 150 LeuLysGlySerSerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePhe 169
Db 260 TTCAGGGTGTACTCGGGGGACCCCTTGTGTGTGAC---AACCAAGCATATGACTTTTC 204
QY 170 ArgAlaAla-----ValSerThrArgGlyValAlaLysAlaValAspPheIle 185
Db 203 GCCTATGCAAAAAACGGAACAATCTCTTCAGGAATCTTCACTAAGTTGTGCACCTCTCTG 144
QY 186 Pro 186
Db 143 CCG 141

RESULT 6
LOCUS BX238988 644 bp DNA linear GSS 29-JAN-2003
DEFINITION Danio rerio genomic clone DKEY-283L13, genomic survey sequence.
ACCESSION BX238988
VERSION BX238988.1 GI:28161322
KEYWORDS GSS.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE Humphray,S.J., Huckle,E. and Durham,J.L.
1 (bases 1 to 644)
Direct Submission
Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Unpublished
This sequence was generated from the T7 end of BAC 283L13. 283L13
is part of the Daniokey BAC Library created by R. Plasterk and N.V.
Keygene. Further details:
http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES
source
1..644
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEY-283L13"
/tissue_type="Testis"
/note="vector pIndigoBAC-536"
BASE COUNT 129 a 212 c 176 g 127 t
ORIGIN
Alignment Scores:

```

```

Pred. No.: 10.7 Length: 644
Score: 97.00 Matches: 48
Percent Similarity: 41.95% Conservative: 25
Best Local Similarity: 27.59% Mismatches: 84
Query Match: 9.51% Indels: 17
DB: Gaps: 7

US-09-965-594-20 (1-197) x BX238988 (1-644)
QY 18 AspThrAlaTyrAlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThr 37
Db 556 GACCATGGCTATTACCATCACTCGAGGGGGAGACAGCTCGAATCCGGACGGCCATCTC 497
QY 38 GlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThr 57
Db 496 CTTGACGGGTGCAGCCAGCCGGAAGCCCTTCCTCGCCGGTGCAGAGTGCATTAGGGCC 437
QY 58 PheLeuAlaThrSerIleAsnGlyValLeu-TripThrValTyrHisGlyAlaGlyThrAr 77
Db 436 GAGACTGCATGGAGACCCCGCGTCCGACAGCTGCGACAGCATCTCGCAGAGTGGGAGCTC 377
QY 77 gThrIleAlaSerProLysGlyProValThrGlnMetTyr-----ThrAs 92
Db 376 TCGAGGCGCTGTCCGCGCTCACAGTGCACGACGACGCTGGTCATGTGGTGTGCCAGAG 317
QY 92 nValAspLysAspLeuValGlyTyrPlnAla----ProGlnGlySerArg-----Se 108
Db 316 TCITGGAAAGGAC-----TGGAGATCCACACCGTTGGGGACAAAGCGGAGAGTCTC 266
QY 108 rLeuThrProCysThrCysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspVa 128
Db 265 GTCACCAACCATGCCCTTTAAAAAAGGAGAAATATTATTGATTACCATGGGAGGAGTTG 206
QY 128 lIleProValArgArgGlyAspSerArgGlySerLeuLeuSerProArgPro----- 146
Db 205 TCGCAGAGAGGGCTTAGGGAGCGGAGCGCTCCCGCTGTCTCCCTCTACATG 146
QY 147 -IleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuLeu---CysProAlaGlyHisAl 165
Db 145 TATTCTCT--TTAAAGGCTTGTGGAGGACCCCTTTCCTGGATGCCAGTCTCGGCC 89
QY 165 aValGlyIlePheArgAlaAlaValSerThrArgGlyVal 178
Db 88 TGCCCTCTGCACCCGGGCATGGAACCTTCGGAAGGCTTA 49

RESULT 7
LOCUS BF863244 701 bp mRNA linear EST 19-JAN-2001
DEFINITION 963042C02.x1 C. reinhardtii CC-1690, Stress condition 1, normalized
, Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION BF863244
VERSION BF863244.1 GI:12253388
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadales; Chlamydomonas.
REFERENCE 1 (bases 1 to 701)
AUTHORS Grossman,A., Davies,J., Federspiel,N., Harris,E., Hauser,C.,
Lefebvre,P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.
Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants; project phase 3
Unpublished
Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.
FEATURES
source
1..701

```

/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="CC-1690 wild type mt+ 2lgr"
/db_xref="taxon:3055"

/clone_lib="c. reinhardtii CC-1690, Stress condition I, normalized, Lambda Zap II"
/note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min, 1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr, 4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda Zap clones by superinfection with EXAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT 173 a 213 c 175 g 140 t

Alignment Scores:
Pred. No.: 16.8 Length: 701
Score: 95.50 Matches: 38
Percent Similarity: 38.96% Conservatives: 22
Best Local Similarity: 24.68% Mismatches: 63
Query Match: 9.36% Indels: 31
DB: 10 Gaps: 7

US-09-965-594-20 (1-197) x BF863244 (1-701)

QY 71 TyrHisGlyAlaGlyThrArgThrIleAlaSerProLys-----GlyProVal 86
Db 171 CACCACCATACCCCTGCTCTCAGCTGCTCACACCAAAATATGCCATACGGGGCACTA 230
QY 87 ThrGlnMetTyrrAsnValAspLysAspLeuValGlyTrpGlnAlaProGlnGlySer 106
Db 231 ACAAGTTTACATACACGG-----AAGCACCAGCGCGCTGGCCACCCCTTGGAGCGG 284
QY 107 ArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrrLeuValThrArgHisAla 126
Db 285 AGAAGCCGACCGGTGCTCTGGGTGTCATCCGCATGCTATGCAATCTCCCGCTATCAG 344
QY 127 AspValIle-----ProValArgArgArgGlyAspSerArg----- 138
Db 345 GAGATCATTTGTCATGCTGGCTTTAGTCACCCCAAGAGAGCGCTGGAGTGGCATTTATAA 404
QY 139 -----GlySerLeuSerProArgProLysSer---Tyr 149
Db 405 GAAGGGNCGGAATTCGTTTTCGGAAGAGTGAAGCCCGCCCAAGGCTGACCAAGTGCTA 464
QY 150 LeuLysGlySerSerGlyProLeuLeuLysProAlaGlyHisAlaValGlyIlePhe 169
Db 465 CTCGAAGCGAGCAATGGGAGCGTTTCGCGGTGTCGCGGTGCTGCTCCCTCTAATGTGTCAGG 524
QY 170 ArgAlaAlaVal-----SerThrArgGlyValAlaLysAla--- 181
Db 525 AAAGAACCATTGAGTAGGAAGTGCAGCGCTTTACCCCGCAAGGTGAAGTCACTCTAT 584
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArg 195
Db 585 GTGACCCGCATACATCTGGAAGACACACGCGGTGCACICTACGA 626

RESULT 8
BF304699/c 984 bp mRNA linear EST 21-NOV-2000
LOCUS 601888252F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:412276 5',
DEFINITION mRNA sequence.
ACCESSION BF304699
VERSION BF304699.1 GI:11251586
KEYWORDS EST.
SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 984)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgaabs-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov

Plate: LLCM1005 row: g column: 13

High quality sequence stop: 646.

Location/Qualifiers

1. .984

source

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:412276"

/tissue_type="rhabdomyosarcoma"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH_MGC_17"

/note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI;

Site_2: XhoI; cDNA made by oligo-dT priming.

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGCAGCAG(G). Size-selected >500bp

for average insert size 1.8kb. Library constructed by

Ling Hong in the laboratory of Gerald M. Rubin (University

of California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 133 a 329 c 351 g 171 t

ORIGIN

Alignment Scores:
Pred. No.: 26.2 Length: 984
Score: 95.50 Matches: 39
Percent Similarity: 42.86% Conservatives: 6

Best Local Similarity: 37.14% Mismatches: 26

Query Match: 9.36% Indels: 34

DB: 10 Gaps: 7

US-09-965-594-20 (1-197) x BF304699 (1-984)

QY 100 TrpGlnAlaProGlnGlySerArgSerLeuThr---ProCysThrCysGlySerSerAsp 118
Db 646 TGGCCCACTCCACGCAATCCCGTGGCAGAGAGACCGGTGTACCTGC----- 599
QY 119 LeuTyrrLeuValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArg 138
Db 598 -----ACCAGCAGCAGCGCAACATACATCAAGAGACGTGT---TCCCGC 554
QY 139 GlySerLeuLeuSerProArgProLysSerTyrrLeuLysGlySerGlyGlyProLeu 158
Db 553 GGGCCCTCTTG-----TGGGAGACCTCGA 527
QY 159 LeuCysProAlaGlyHis-Ala-----ValGlyIlePheArg-----AlaAlaVal 173
Db 526 TGGTGTCCAAGCTCGCGTGTGTACTGGAAGTCGGCAGCGCTCCGGTCAGTGCGAGCTTC 467
QY 173 IserThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrTh 193
Db 466 CAGCCCGCGGGTGGC-----CCGAGGAGCTCTCAGGGTCCC 428
QY 193 rMetArgSerPro 197
Db 427 CAAGGGGGCGGCC 415

RESULT 9

AQ538021

Qy 18 AspThrAlaTyraAlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThr 37
 Db 611 GCCAAACATACACGGACCAACAGATACGGGTACAGGGTGCCAAAG----- 658
 Qy 38 GlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThr 57
 Db 659 GGAGCGGATAAAGGCTAGGCAACGGG-----CCCCGAGTAACCC 700
 Qy 58 PheLeuAlaThrSerIleAsnGlyValLeuTyrThrValTyrHisGlyAlaGlyThrArg 77
 Db 701 GCGTCGGCGCGGAGGACACGGGAAACCCCGGGAAACCCCTCGGAGCAGGGAACGCTG 760
 Qy 78 ThrIleAlaSerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeu 97
 Db 761 ACAAAACCGCGGAGGAGGACCTCAACGCGGCCCCACACCAACCTGACCGCAACAT- 819
 Qy 98 ValGlyTyrGlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSer 117
 Db 820 -----TACACGCCCCACACGGG-----ACACCACTA----- 846
 Qy 118 AspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgGlyAspSer 137
 Db 847 CACAAATTTCAACAGCGCGGCTACCCCGGGTCCACCTACGCGGAGAAATCATCAGCGC 906
 Qy 138 ArgGlySer 140
 Db 907 CGAGCGTCA 915

RESULT 11

CNS06QHN 1062 bp DNA linear GSS 05-JUL-2001
 LOCUS T3 end of clone AWOAA006803 of library AWOAA from strain CLIB 89 of
 Yarrowia lipolytica, genomic survey sequence.
 ACCESSION AL410673
 VERSION AL410673.1 GI:12179275
 KEYWORDS GSS.
 SOURCE Yarrowia lipolytica
 ORGANISM Yarrowia lipolytica
 Eukaryote; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Dipodascaceae; Yarrowia.
 REFERENCE 1 (bases 1 to 1062)
 AUTHORS Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
 Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
 de-Montigny,J., DuJon,B., Durrens,P., Lepingle,A., Llorente,B.,
 Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
 Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
 Wincker,P. and Weissbach,J.
 TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of
 yeast species for molecular evolution studies
 JOURNAL FEMS Lett. 487 (1), 3-12 (2000)
 MEDLINE 20584711
 PUBMED 11152876
 REFERENCE 2 (bases 1 to 1062)
 AUTHORS Casaregola,S., Neuveglise,C., Lepingle,A., Bon,E., Feynerol,C.,
 Artiguenave,F., Wincker,P. and Gaillardin,C.
 TITLE Genomic exploration of the hemiascomycetous yeasts: 17. Yarrowia
 lipolytica
 JOURNAL FEMS Lett. 487 (1), 95-100 (2000)
 MEDLINE 20584727
 PUBMED 11152892
 REFERENCE 3 (bases 1 to 1062)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
 2 rue Gaston Cremieux, Cp 5706, 91057 EVRY cedex, FRANCE. (E-mail :
 seqrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)
 COMMENT This GSS is part of a random genomic sequencing program of thirteen
 yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
 exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,
 Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
 lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
 angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
 Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to

5 kb were prepared and both extremities were sequenced. See
 keywords for description of this sequence and for the sequence of
 the other extremity of this insert.

FEATURES

source
 1..1062
 Location/Qualifiers
 /organism="Yarrowia lipolytica"
 /mol_type="genomic DNA"
 /strain="CLIB 89"
 /db_xref="taxon:4952"
 /clone="AWOAA006803"
 /clone_lib="AWOAA"
 /note="end : T3"
 <594..>917
 /note="similar to Saccharomyces cerevisiae ORF YKR093w [
 PTR2 : peptide transporter]"
 /evidence=not_experimental

misc_feature

BASE COUNT 274 a 264 c 242 g 280 t 2 others
 ORIGIN

Alignment Scores:

Pred. No.: 40.8 Length: 1062
 Score: 94.00 Matches: 50
 Percent Similarity: 36.10% Conservative: 24
 Best Local Similarity: 24.39% Mismatches: 57
 Query Match: 9.22% Indels: 74
 DB: 29 Gaps: 11

US-09-965-594-20 (1-197) x CNS06QHN (1-1062)

Qy 32 GlnLysThrSerHisThr-----GlyArgAspLysAsnGlnValGlu---GlyGluVal 48
 Db 447 CAAGAAGCTACCCACTCTATCGATAAGACGACAAAAGAAATGAATTCACCGAGATT 506
 Qy 49 GlnIleValSerThrAlaThrGlnThrPheLeuAlaThrSerIleAsnGlyValLeuTyr 68
 Db 507 GAACACATCGATCATCTCCCGGGATACTCTCAGACA-----TGG 548
 Qy 69 ThrValTyr-----HisGly----- 73
 Db 549 GCCACTACACCGATGAACATAACCCGACGGCTCAGAATTGCCACAGAGAAATCGA 608
 Qy 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValThrGlnMet----- 89
 Db 609 GCGACTTAAGCAGGATAGCTGCTCCAAATTGGAGCCATGACTTACATGCTCTCTTGT 668
 Qy 90 -----TyrThrAsnValAspLysAspLeuValcyl 99
 Db 669 GAGTTGCGCAGAGCGAGGCTCGTACTATGCTGACCAACGTCATTTCACACTTTGTCAG 728
 Qy 100 TrpGlnAlaProGlnGlySerArg-----SerLeuThrProCys-----Thr 113
 Db 729 TTCCCTCTCCCTAAGGCGGAAATGGCTGGGAGCCACACCTCGCGGTTCACATTGACA 788
 Qy 114 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArg 133
 Db 789 GCGGAGCCCTAGATCAGGGTCTTCAAGTCGCCAGGCTCTGACTCTGGTTTTCAGTTT 848
 Qy 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 153
 Db 849 TTA-----TCCTACTTGACCTCTTCTAGGAGCCCTACCTTCCCGATTC 893
 Qy 154 Ser-----GlyGlyProLeuLeuCysProAlaGlyHis--- 164
 Db 894 AAATATGACGTTTCAAGACCATCTGGGCCGCTAGCATCTCTGGNAATGGCCATATT 953
 Qy 165 -----AlaValGlyIle 168
 Db 954 GTGATTGTGATTCGCGGAATTCGCCGAATCATAGACCAAGAAAGCTCTCTAGGAATC 1013
 Qy 169 PheArgAlaAlaVal 173
 Db 1014 TTTATTGCTGGACTA 1028

RESULT 12

```

CC406704/c
LOCUS          CC406704 772 bp DNA linear GSS 19-MAY-2003
DEFINITION    PUHKL12TD 2M_0.6_1.0_KB zea mays genomic clone ZMMBTa469B24,
               genomic survey sequence.
ACCESSION     CC406704
VERSION       CC406704.1 GI:30886794
KEYWORDS      GSS.
SOURCE        Zea mays
ORGANISM      Zea mays
               Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
               Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
               clade; Panicoideae; Andropogoneae; Zea.
REFERENCE     1 (bases 1 to 772)
AUTHORS       Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick
               A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.
TITLE         Maize Genomics Consortium
JOURNAL       Unpublished
COMMENT       Other_GSSs: PUHKL12TD
               Contact: Cathy Whitelaw
               TIGR
               9712 Medical Center Drive, Rockville, MD 20850, USA
               Tel: 301-838-5843
               Fax: 301-838-0208
               Email: whitelaw@tigr.org
               Seq primer: TR
               Class: sheared ends.
               Location/Qualifiers
                 1..772
                   /organism="Zea mays"
                   /mol_type="genomic DNA"
                   /strain="B73"
                   /db_xref="taxon:4577"
                   /clone_lib="ZMMBTa469B24"
                   /clone_lib="2M_0.6_1.0_KB"
                   /note="vector: PCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                   Cot selected genomic DNA library"
BASE COUNT    150 a 198 c 216 g 208 t
ORIGIN
Alignment Scores:
Pred. No.:      30.1 Length:      772
Score:          93.50 Matches:      49
Percent Similarity: 39.46% Conservative: 24
Best Local Similarity: 26.49% Mismatches: 67
Query Match:     9.17% Indels:      46
DB:              29 Gaps:          10

US-09-965-594-20 (1-197) x CC406704 (1-772)

Qy 29 GlnGlyCysGlnLysThrSerHisThrGlyArgAspLysAsnGlnVal---GluGlyGlu 47
Db 663 CAGTCCTGTTCAAGCGCCTCACCCCGCGAGATGGCGGAACGACGTAACAAGGGCTAT 604
Qy 48 ValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThrSerIleAsnGlyVal--- 66
Db 603 GTTACAAATGTGAGC-----AGCCCTATGTACGAGGCCCGCTGTCCAGGCTATTCT 550
Qy 67 LeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLysGlyProVal 86
Db 549 ATCTGGAGG-----TCACAGACATTTGCTGACGATGAGATCAAGGGTG 508
Qy 87 ThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp-----GlnAlaPro 103
Db 507 ACACATGATGATGACAGG-----AAGAGCAGGAGCCAGTGTTCCTTACATGCAATAA 454
Qy 104 GlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeu----- 119
Db 453 CTGGGATTAGAGGGAGGACCATCAGCTGGCGGTAGTCTCTCAACGGTCAAGGAGCTGC 394
Qy 120 -----TyrLeuValThrArgHisAlaAspValIleProValArgArgGlyAspSer 137
Db 393 TGGCCCTTACTTGACGGGTTCACACATAACTTCATCACTCAAGGGGGCA-CNACAG 335

```

```

Qy 138 ArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyPro 157
Db 334 CGTGGGTACTTGGAAACCCACACAGGTCGCATGTCGAAGTGGCAATATGGAGACCCA 275
Qy 158 LeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGly 177
Db 274 GTTTTCTGCCAG-----GGAGTAACTCGTCGCGCA----- 245
Qy 178 ValAlaLysAlaValAspPhe-----IlePro 186
Db 244 -----GCCATTGACATCAACAGGAGAGTTCACCATGAGGCATATGCAATTCCC 194
Qy 187 ValGluSerLeuGlu 191
Db 193 TTGATACATTGAG 179

RESULT 13
LOCUS          CC406705
DEFINITION    PUHKL12TD 2M_0.6_1.0_KB zea mays genomic clone ZMMBTa469B24,
               genomic survey sequence.
ACCESSION     CC406705
VERSION       CC406705.1 GI:30886795
KEYWORDS      GSS.
SOURCE        Zea mays
ORGANISM      Zea mays
               Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
               Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
               clade; Panicoideae; Andropogoneae; Zea.
REFERENCE     1 (bases 1 to 789)
AUTHORS       Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick
               A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.
TITLE         Maize Genomics Consortium
JOURNAL       Unpublished
COMMENT       Other_GSSs: PUHKL12TB
               Contact: Cathy Whitelaw
               TIGR
               9712 Medical Center Drive, Rockville, MD 20850, USA
               Tel: 301-838-5843
               Fax: 301-838-0208
               Email: whitelaw@tigr.org
               Seq primer: TF
               Class: sheared ends.
               Location/Qualifiers
                 1..789
                   /organism="Zea mays"
                   /mol_type="genomic DNA"
                   /strain="B73"
                   /db_xref="taxon:4577"
                   /clone_lib="ZMMBTa469B24"
                   /clone_lib="2M_0.6_1.0_KB"
                   /note="vector: PCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                   Cot selected genomic DNA library"
BASE COUNT    210 a 220 c 203 g 156 t
ORIGIN
Alignment Scores:
Pred. No.:      31 Length:      789
Score:          93.50 Matches:      49
Percent Similarity: 39.46% Conservative: 24
Best Local Similarity: 26.49% Mismatches: 67
Query Match:     9.17% Indels:      46
DB:              29 Gaps:          10

US-09-965-594-20 (1-197) x CC406705 (1-789)

Qy 29 GlnGlyCysGlnLysThrSerHisThrGlyArgAspLysAsnGlnVal---GluGlyGlu 47
Db 110 CAGTCTGTTCAAGCGCCTCACCCCGCGAGATGGCGGAACGACGTAACAAGGGCTAT 169
Qy 48 ValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThrSerIleAsnGlyVal--- 66

```

```

Db      170 GTTACAATGTGACG-----AGCCCTATGTACGAGGCCACCGCTGCCAGGCTATTCT 223
QY      67 LeuTprThrValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLysGlyProVal 86
Db      224 ATCTGGAG-----TCACAGACTTGTGACGATGACGATCAAGGTTG 265
QY      87 ThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr-----GlnAlaPro 103
Db      266 ACACAGATGATGACCAGG-----AAGAGCAGGAGCCAGTGGTTTCCCTACATGCAATAA 319
QY      104 GlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeu----- 119
Db      320 CTGGGATTAGAGGAGGAGGAGCACCACGCTGCGGGTAGTCTCAACGGTCAGGAGTGTC 379
QY      120 -----TyrLeuValThrArgHisAlaAspValIleProValArgArgGlyAspSer 137
Db      380 TGGCCCTACTTGACACGGTTCACACATACTTCATCACTGCAAGCGGSCA-CAACAG 438
QY      138 ArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyPro 157
Db      439 CTGGGGTTACTTTGGAAACCCACACAGGTCGCCATGTCAAGGTGGCAATGGAGACCCA 498
QY      158 LeuLeuSerProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGly 177
Db      499 GTTTCTGCCAG-----GGAGTACTGTCGCCGA----- 528
QY      178 ValAlaLysAlaValAspPhe-----IlePro 186
Db      529 -----GCCATTGCATCAACCAAGGAGAAGTTCAACCAATTGAGCATATGCAATTC 579
QY      187 ValGluSerLeuGlu 191
Db      580 TTGGATACATTGAG 594

RESULT 14
BM402566
LOCUS      528 bp mRNA linear EST 01-JUL-2002
DEFINITION SLA005f12_34513 An expressed sequence tag (EST) collection from the
            resurrection plant Selaginella lepidophylla
ACCESSION  SLA005f12 5, mRNA sequence.
VERSION     BM402566
KEYWORDS    Selaginella
SOURCE      BM402566.1 GI:21643782
ORGANISM    Selaginella lepidophylla
            Selaginella lepidophylla
            Eukaryota; Viridiplantae;
            Lycopodiophyta; Isoetopsida; Selaginellales; Selaginellaceae;
            Selaginella.
REFERENCE   1 (bases 1 to 528)
AUTHORS    Iturriaga,G. and Cushman,J.C.
TITLE      An expressed sequence tag (EST) collection from the resurrection
JOURNAL    Plant Selaginella lepidophylla
COMMENT    Unpublished
            Contact: Cushman JC
            Department of Biochemistry
            University of Nevada
            MS200, Reno, NV 89557-0014, USA
            Tel: 775-784-1918
            Fax: 775-784-1650
            Email: jcushman@unr.edu
PCR PRIMERS
FORWARD: T3 20mer
BACKWARD: T7 21mer
Plate: 005 row: f column: 12
Seq primer: T3 20mer
High quality sequence stop: 528.
Location/Qualifiers
            1..528
            /organism="Selaginella lepidophylla"
            /mol_type="mRNA"
            /db_xref="taxon:59777"
            /clone="SLA005f12"
            /tissue_type="microphyll fronds undergoing desiccation for
FEATURES
source

```

```

2.5 h"
/dev_stage="adult"
/clone_lib="An expressed sequence tag (EST) collection
from the resurrection plant Selaginella lepidophylla"
/notes-Vector: Lambda Uni-Zap XR, Bluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Library construction was performed
according to manufacture's (Stratagene, Inc.) recommended
protocol for the Lambda UniZapXR vector and cDNA synthesis
kit."
BASE COUNT 129 a 125 c 137 g 137 t
ORIGIN
Alignment Scores:
Pred. No.: 20.6 Length: 528
Score: 93.00 Matches: 37
Percent Similarity: 42.98% Conservative: 15
Best Local Similarity: 30.58% Mismatches: 43
Query Match: 9.12% Indels: 26
DB: 12 Gaps: 4
US-09-965-594-20 (1-197) x BM402566 (1-528)
QY      94 AspLysAspLeuValGlyTyrGlnAlaProGlnGlySerArgSerLeuThrProCysThr 113
Db      53 GACAAGATGTAGCGGTGCTGAAGATCGATGCTCAAGCAACAGATCTCAGCCCAATACCC 112
QY      114 CysGlySerSerAspLeuTyrLeuVal----- 122
Db      113 CTGGGAAGTTCGTCCGATCTGCTTGTGGCCAGAGGTGTATGCTATCGTAAATCCTTTT 172
QY      123 -----ThrArgHisAlaAspValIleProValArgArgGlyAspSerArg 138
Db      173 GGATTGGATCATACGCTGACACACAGCGCTCATCGTGTCTTGAAGGAGAGATTACT--- 229
QY      139 GlySerLeuLeuSerProArgProIleSerTyrLeu----- 150
Db      230 ---TCAGCCGCTATGCTGCTCAATCCAGACAGTATCCACAGAGATGCCGCTATTAT 286
QY      151 LysGlySerSerGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArg 170
Db      287 CTGGGACACAGCGGGTCCGCTATTGGACAGTCTCTGGAAATTTGATAGGCATCAACT 346
QY      171 AlaAlaValSerThrArgGlyValAlaLysAlaValAspPhe---IleProValGluSer 189
Db      347 GCTATATATTCCTCGCTGCGGCTTCATCAGCGGTGGGCTTTTCATCCAGTTGACACG 406
QY      190 Leu 190
Db      407 GTT 409
RESULT 15
BZ342381/c
LOCUS      701 bp DNA linear GSS 06-NOV-2002
DEFINITION ic83b11.b1 WGS-SbicolorF (JM107 adapted methyl filtered) Sorghum
            bicolor genomic clone ic83b11 5', genomic survey sequence.
ACCESSION  BZ342381
VERSION     BZ342381.1 GI:24742983
KEYWORDS    GSS.
SOURCE      Sorghum bicolor (sorghum)
ORGANISM    Sorghum bicolor
            Eukaryota; Viridiplantae;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1 (bases 1 to 701)
AUTHORS    Rabinowicz,P.D., O'Shaughnessy,A.L., Balija,V., Dedhia,N.,
            Katzenburger,F., King,L., Miller,B., Muller,S., Nascimben,L.,
            Zutavern,T., Palmer,L., McCombie,W.R. and Martienssen,R.A.
TITLE      Genomic shotgun sequences from Sorghum bicolor (methyl-filtered)
JOURNAL    Unpublished
COMMENT    Contact: W. Richard McCombie
            Lita Annenberg Hazen Genome Sequencing Center
            Cold Spring Harbor Laboratory
            PO Box 100, Cold Spring Harbor, NY 11724, USA

```

Search completed: August 31, 2003, 04:27:46
Job time : 1915.31 secs

Tel: 516 367 8884
Fax: 516 367 8874
Email: mcombie@cshl.org
Plate: ic83 row: b column: 11
Seq primer: -21M13UnivFwd
Class: shotgun
High quality sequence stop: 701.
Location/Qualifiers
1. 701

FEATURES

source

/organism="Sorghum bicolor"
/mol_type="genomic DNA"
/db_xref="taxon:4558"
/clone="ic83b1"
/lab_host="JM107 or DH5a"
/clone_lib="WGS-SbicolorF (JM107 adapted methyl filtered)"
/note="Site 1: Xba 1; Site 2: Xba 1; The vector was digested with Xba1 and one nucleotide was added by fill in in the recessive 3' end. The genomic DNA was nebulized, end repaired, adaptor ligated and size fractionated using sephadex. The resulting fragments were between 0.8 and 3 kb and were cloned into the vector (x/y reads in M13mp19, -b/g reads in pUC19). The same ligation was transformed in either JM107 or DH5a."

BASE COUNT 108 a 251 c 232 g 110 t
ORIGIN

Alignment Scores:

Pred. No.:	29.8	Length:	701
Score:	93.00	Matches:	48
Percent Similarity:	36.69%	Conservative:	14
Best Local Similarity:	28.40%	Mismatches:	56
Query Match:	9.12%	Indels:	51
DB:	29	Gaps:	8

US-09-965-594-20 (1-197) x BZ342381 (1-701)

QY	60	AlaThrSerIleAsnGlyValLeuThrThrValTyrHisGlyAlaGlyThrArgThr---	78
DB	500	GCACATGCCGTGCGGGTGTGTTTTGAACGTACCGCACGCGCGGGGAGGACGACAGCC	441
QY	79	-----IleAlaSerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAsp	96
DB	440	GTGGCGGTGGCG	381
QY	97	LeuValGlyTrpGlnAlaProGlnGlySerArg-----SerLeuThrProCysThr---	113
DB	380	CTACGCGGAGCGCGGAGCAGGAGGTGCGCGCTGCTCTCTCCCTCCCTGCTGCTGCTC	321
QY	114	-----CysGlySerSerAsp	118
DB	320	CACCCAGTGCCCG	261
QY	119	LeuTyrLeuValThrArgHisAlaAspValIleProValArgArgGlyAsp-----	136
DB	260	GTGGAGAGGCGGAGCAGCAGCAGCA-----CGACGACGGGTGGAGACCCG	216
QY	137	-----SerArgGlySerLeuLeuSerProArgProIleSerTyr	149
DB	215	GTTTGACG	156
QY	150	LeuLysGly-----SerSerGlyGlyProLeu-LeuCy	160
DB	155	TTTACTGCTCGCCCTCG	96
QY	160	sProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyVal-----	178
DB	95	CGTGGCAGTGGCAGCGGCT-----CGTACGTGAGTACCGCGGAGTCATCAG	48
QY	179	-AlaLysAlaValAspPheIlePro	186
DB	47	TGCCCTCGCGCTTGGTGCCTGCT	23

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 17:42:58 ; Search time 44.6227 Seconds
(without alignments)
700.745 Million cell updates/sec

Title: US-09-965-594-22

Perfect score: 1016

Sequence: 1 MKKKGWVIVGRINLSGDTA.....YAKAVDFIPVESLETTMRSP 197

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

```

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1016	100.0	197	AA15225	Hepatitis C virus
2	1010	99.4	197	AA15224	Hepatitis C virus
3	995	97.9	197	AA15223	Hepatitis C virus
4	980	96.5	197	AA15222	Hepatitis C virus
5	963	94.8	197	AA15221	Hepatitis C virus
6	936	92.1	197	AA15226	Hepatitis C virus
7	929	91.4	195	AA15220	Hepatitis C virus
8	902	88.8	195	AA15212	Hepatitis C virus
9	875.5	86.2	665	AA124943	HCV NS4A-NS3 compl

10	872.5	85.9	665	20	AA124947	HCV NS4A-NS3 compl
11	871.5	85.8	665	20	AA124942	HCV NS4A-NS3 compl
12	868.5	85.5	216	20	AA17880	HCV NS4A-NS3 compl
13	868.5	85.5	665	20	AA124946	HCV NS4A-NS3 compl
14	867.5	85.4	665	20	AA124941	HCV NS4A-NS3 compl
15	865.5	85.2	216	20	AA17884	HCV NS4A-NS3 compl
16	864.5	85.1	216	20	AA17879	HCV NS4A-NS3 compl
17	864.5	85.1	665	20	AA124945	HCV NS4A-NS3 compl
18	863.5	85.0	665	20	AA124940	HCV NS4A-NS3 compl
19	863.5	85.0	671	20	AA124948	HCV NS4A-NS3 compl
20	861.5	84.8	216	20	AA17883	HCV NS4A-NS3 compl
21	860.5	84.7	216	20	AA17878	HCV NS4A-NS3 compl
22	860.5	84.7	665	20	AA124944	HCV NS4A-NS3 compl
23	860.5	84.7	671	20	AA124949	HCV NS4A-NS3 compl
24	860	84.6	215	20	AA17890	HCV NS4A-NS3 compl
25	857.5	84.4	216	20	AA17882	HCV NS4A-NS3 compl
26	857.5	84.4	216	20	AA17886	HCV NS4A-NS3 compl
27	856.5	84.3	216	20	AA17877	HCV NS4A-NS3 compl
28	854	84.1	215	20	AA17887	HCV NS4A-NS3 compl
29	853.5	84.0	216	20	AA17881	HCV NS4A-NS3 compl
30	853.5	84.0	216	20	AA17885	HCV NS4A-NS3 compl
31	849	83.6	213	20	AA17888	HCV NS4A-NS3 compl
32	849	83.6	631	20	AA17882	HCV NS3 protein..
33	848.5	83.5	131	21	AA144728	Hepatitis C virus
34	848.5	83.5	3011	19	AA177397	Hepatitis C virus
35	848.5	83.5	3011	24	ABP71460	Amino acid sequence
36	848.5	83.5	3012	23	AA199289	Hepatitis C virus
37	845.5	83.2	3011	14	AA140120	HCV genomic amino
38	844.5	83.1	687	16	AA179223	PHCV150-encoded se
39	844.5	83.1	1648	16	AA179221	PHCV176-encoded se
40	844.5	83.1	1766	10	AA192041	Sequence encoded i
41	844.5	83.1	1766	10	AA192041	Protein sequence o
42	844.5	83.1	2261	10	AA190164	Peptide encoded by
43	844.5	83.1	2301	10	AA192047	Sequence encoded i
44	844.5	83.1	2436	10	AA192050	Sequence encoded i
45	844.5	83.1	2436	10	AA192088	Peptide encoded by

ALIGNMENTS

RESULT 1
AAB15225
ID AAB15225 standard; protein: 197 AA.
XX AC AAB15225;
XX DT 19-DEC-2000 (first entry)
XX DE Hepatitis C virus NS4A-NS3 fusion protease #7.
XX KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
XX KW liver failure; liver cancer; mutant; mutain.
XX OS Hepatitis C virus.
XX OS Synthetic.
XX PN WO200040707-A1.
XX PD 13-JUL-2000.
XX PF 06-JAN-2000; 2000WO-US00345.
XX PR 08-JAN-1999; 99US-0115271.
XX PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX DR WPI; 2000-465976/40.
XX DR N-PSDB; AAA73334.
XX PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1

PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 PT -
 XX
 PS Claim 23; Fig 17; 66pp; English.
 XX
 CC The present sequence is a mutated version of a fusion protein created
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
 CC proteins are both essential for the replication of the virus, acting to
 CC cleave its replicative proteins from the polyprotein produced from the
 CC HCV genome. Inhibitors of the two proteins should be effective as
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A
 CC fusion proteins which can be used to identify inhibitors of this type, as
 CC well as enabling structural studies of the protease and
 CC protease:inhibitor complexes. This sequence contains the alpha-helix0-7
 CC variant.
 XX
 SQ Sequence 197 AA;

Query Match 100.0%; Score 1016; DB 21; Length 197;
 Best Local Similarity 100.0%; Pred. No. 2.1e-98;
 Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKKGSVVIVGRINLSGDTAYAAQTGREGQGTQKTSHTGRDKNQVEGEVQIVSTATQTFLA 60
 DB 1 MKKGSVVIVGRINLSGDTAYAAQTGREGQGTQKTSHTGRDKNQVEGEVQIVSTATQTFLA 60
 QY 61 TSINGVLWTVYHGAGTRTIASPKGPVTQMTYNDKDLVGHQAPQGSRLTPTCTCGSSDLY 120
 DB 61 TSINGVLWTVYHGAGTRTIASPKGPVTQMTYNDKDLVGHQAPQGSRLTPTCTCGSSDLY 120
 QY 121 LVTRHADVIPVRRRGRSGLSLSPRISYLYKSGSGPLLCPCPAGHAGVIFRAAVSTRGVAK 180
 DB 121 LVTRHADVIPVRRRGRSGLSLSPRISYLYKSGSGPLLCPCPAGHAGVIFRAAVSTRGVAK 180
 QY 181 AVDFIPVESLETTMRSP 197
 DB 181 AVDFIPVESLETTMRSP 197

RESULT 2
 AAB15224
 ID AAB15224 standard; protein; 197 AA.
 AC AAB15224;
 XX
 XX
 DT 19-DEC-2000 (first entry)
 XX
 DE Hepatitis C virus NS4A-NS3 fusion protease #6.
 XX
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
 KW liver failure; liver cancer; mutant; mutein.
 OS Hepatitis C virus.
 OS Synthetic.
 XX
 PN WO200040707-A1.
 XX
 PD 13-JUL-2000.
 XX
 PF 06-JAN-2000; 2000WO-US00345.
 XX
 PR 08-JAN-1999; 99US-0115271.
 XX
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX
 PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
 XX
 DR WPI: 2000-465976/40.
 DR N-PSDB; AAA73333.
 XX

PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 PT -
 XX
 PS Claim 23; Fig 16; 66pp; English.
 XX
 CC The present sequence is a mutated version of a fusion protein created
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
 CC proteins are both essential for the replication of the virus, acting to
 CC cleave its replicative proteins from the polyprotein produced from the
 CC HCV genome. Inhibitors of the two proteins should be effective as
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A
 CC fusion proteins which can be used to identify inhibitors of this type, as
 CC well as enabling structural studies of the protease and
 CC protease:inhibitor complexes. This sequence contains the alpha-helix0-7
 CC variant.
 XX
 SQ Sequence 197 AA;

Query Match 99.4%; Score 1010; DB 21; Length 197;
 Best Local Similarity 99.5%; Pred. No. 9e-98;
 Matches 196; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MKKGSVVIVGRINLSGDTAYAAQTGREGQGTQKTSHTGRDKNQVEGEVQIVSTATQTFLA 60
 DB 1 MKKGSVVIVGRINLSGDTAYAAQTGREGQGTQKTSHTGRDKNQVEGEVQIVSTATQTFLA 60
 QY 61 TSINGVLWTVYHGAGTRTIASPKGPVTQMTYNDKDLVGHQAPQGSRLTPTCTCGSSDLY 120
 DB 61 TSINGVLWTVYHGAGTRTIASPKGPVTQMTYNDKDLVGHQAPQGSRLTPTCTCGSSDLY 120
 QY 121 LVTRHADVIPVRRRGRSGLSLSPRISYLYKSGSGPLLCPCPAGHAGVIFRAAVSTRGVAK 180
 DB 121 LVTRHADVIPVRRRGRSGLSLSPRISYLYKSGSGPLLCPCPAGHAGVIFRAAVSTRGVAK 180
 QY 181 AVDFIPVESLETTMRSP 197
 DB 181 AVDFIPVESLETTMRSP 197

RESULT 3
 AAB15223
 ID AAB15223 standard; protein; 197 AA.
 AC AAB15223;
 XX
 XX
 DT 19-DEC-2000 (first entry)
 XX
 DE Hepatitis C virus NS4A-NS3 fusion protease #5.
 XX
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
 KW liver failure; liver cancer; mutant; mutein.
 OS Hepatitis C virus.
 OS Synthetic.
 XX
 PN WO200040707-A1.
 XX
 PD 13-JUL-2000.
 XX
 PF 06-JAN-2000; 2000WO-US00345.
 XX
 PR 08-JAN-1999; 99US-0115271.
 XX
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX
 PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
 XX
 DR WPI: 2000-465976/40.
 DR N-PSDB; AAA73332.
 XX

XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
PT amino acid, useful for screening inhibitors that may treat hepatitis C
PT
XX
PS Claim 23; Fig 15; 66pp; English.
XX
CC The present sequence is a mutated version of a fusion protein created
CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
CC proteins are both essential for the replication of the virus, acting to
CC cleave its replicative proteins from the polyprotein produced from the
CC HCV genome. Inhibitors of the two proteins should be effective as
CC antiviral treatments of HCV infection. This is useful as HCV can lead to
CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
CC The present invention concerns a number of NS3 mutants and NS3-NS4A
CC fusion proteins which can be used to identify inhibitors of this type, as
CC well as enabling structural studies of the protease and
CC protease:inhibitor complexes. This sequence contains the alpha-helix0-1
CC variant.
XX
SQ Sequence 197 AA:
Query Match 97.9%; Score 995; DB 21; Length 197;
Best Local Similarity 98.0%; Pred. No. 3.4e-96;
Matches 193; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 1 MKKGSVIVGRINLSGDTAYAQOTRGEQGTOKTSHTGRDKNQVGEQIVSTATQTFLA 60
DB 1 MKKGSVIVGRINLSGDTAYAQOTRGEQGTOKTSHTGRDKNQVGEQIVSTATQTFLA 60
QY 61 TSINGVLWTVYHGAGTTRTIASPKGPVTOMYTNVDKDLVGMQAPQGSRSLLTPCTCGSSDLY 120
DB 61 TSINGVLWTVYHGAGTTRTIASPKGPVTOMYTNVDKDLVGMQAPQGSRSLLTPCTCGSSDLY 120
QY 121 LVTRHADVIPVRRGDSRGSLLSPRISYLGKSSGGPILCPAGHAGVIFRAAVSTRGVAK 180
DB 121 LVTRHADVIPVRRGDSRGSLLSPRISYLGKSSGGPILCPAGHAGVIFRAAVSTRGVAK 180
QY 181 AVDFIPVESLETTMRSP 197
DB 181 AVDFIPVESLETTMRSP 197
RESULT 4
AAB15222
ID AAB15222 standard; protein; 197 AA.
XX
AC AAB15222;
XX
DT 19-DEC-2000 (first entry)
XX
DE Hepatitis C virus NS4A-NS3 fusion protease #4.
XX
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW liver failure; liver cancer; mutant; mutein.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
PN WO200040707-A1.
XX
PD 13-JUL-2000.
XX
PF 06-JAN-2000; 2000WO-US00345.
XX
PR 08-JAN-1999; 99US-0115271.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX WPI; 2000-465976/40.

DR N-PSDB; AAA73331.
XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
PT amino acid, useful for screening inhibitors that may treat hepatitis C
PT
XX
PS Claim 23; Fig 14; 66pp; English.
XX
CC The present sequence is a mutated version of a fusion protein created
CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
CC proteins are both essential for the replication of the virus, acting to
CC cleave its replicative proteins from the polyprotein produced from the
CC HCV genome. Inhibitors of the two proteins should be effective as
CC antiviral treatments of HCV infection. This is useful as HCV can lead to
CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
CC The present invention concerns a number of NS3 mutants and NS3-NS4A
CC fusion proteins which can be used to identify inhibitors of this type, as
CC well as enabling structural studies of the protease and
CC protease:inhibitor complexes. This sequence contains the alpha-helix0-1
CC variant.
XX
SQ Sequence 197 AA:
Query Match 96.5%; Score 980; DB 21; Length 197;
Best Local Similarity 96.4%; Pred. No. 1.3e-94;
Matches 190; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 1 MKKGSVIVGRINLSGDTAYAQOTRGEQGTOKTSHTGRDKNQVGEQIVSTATQTFLA 60
DB 1 MKKGSVIVGRINLSGDTAYAQOTRGEQGTOKTSHTGRDKNQVGEQIVSTATQTFLA 60
QY 61 TSINGVLWTVYHGAGTTRTIASPKGPVTOMYTNVDKDLVGMQAPQGSRSLLTPCTCGSSDLY 120
DB 61 TSINGVLWTVYHGAGTTRTIASPKGPVTOMYTNVDKDLVGMQAPQGSRSLLTPCTCGSSDLY 120
QY 121 LVTRHADVIPVRRGDSRGSLLSPRISYLGKSSGGPILCPAGHAGVIFRAAVSTRGVAK 180
DB 121 LVTRHADVIPVRRGDSRGSLLSPRISYLGKSSGGPILCPAGHAGVIFRAAVSTRGVAK 180
QY 181 AVDFIPVESLETTMRSP 197
DB 181 AVDFIPVESLETTMRSP 197
RESULT 5
AAB15221
ID AAB15221 standard; protein; 197 AA.
XX
AC AAB15221;
XX
DT 19-DEC-2000 (first entry)
XX
DE Hepatitis C virus NS4A-NS3 fusion protease #3.
XX
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW liver failure; liver cancer; mutant; mutein.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
PN WO200040707-A1.
XX
PD 13-JUL-2000.
XX
PF 06-JAN-2000; 2000WO-US00345.
XX
PR 08-JAN-1999; 99US-0115271.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX

DR WPI: 2000-465976/40.
 DR N-PSDB; AAA73330.
 XX
 PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 PT
 XX
 PS Claim 23; Fig 13; 66pp; English.
 XX
 CC The present sequence is a mutated version of a fusion protein created
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
 CC proteins are both essential for the replication of the virus, acting to
 CC cleave its replicative proteins from the polyprotein produced from the
 CC HCV genome. Inhibitors of the two proteins should be effective as
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A
 CC fusion proteins which can be used to identify inhibitors of this type, as
 CC well as enabling structural studies of the protease and
 CC protease:inhibitor complexes. This sequence contains the alpha-helix0-1
 CC variant.
 XX
 SQ Sequence 197 AA;

Query Match 94.8%; Score 963; DB 21; Length 197;
 Best Local Similarity 94.9%; Pred. No. 7. Be-93;
 Matches 187; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 MKKGSVVIVGRINLSGDTAYAOQTREGCGTQKTSHTGRDNKNOVEGEVQIVSTATQTEFLA 60
 DB 1 MKKGSVVIVGRINLSGDTAYAOQTREGCGCQETQSGTGRDNKNOVEGEVQIVSTAAQTFLA 60
 QY 61 TSINGVLWTVYHGAGTRTASPKGPVTQMTYNDKDLVGMQAPQGSRSLSLTPCTCGSSDLY 120
 DB 61 TCINGVCTVYHGAGTRTASPKGPVTQMTYNDKDLVGMQAPQGSRSLSLTPCTCGSSDLY 120
 QY 121 LVTRHADVIPVRRGDSRGLSPRPISYLYKSGSGGPLLCPCPAGHAGVIFRAAVSTRGVAK 180
 DB 121 LVTRHADVIPVRRGDSRGLSPRPISYLYKSGSGGPLLCPCPAGHAGVIFRAAVSTRGVAK 180
 QY 181 AVDFIPVESLETTMRSP 197
 DB 181 AVDFIPVESLETTMRSP 197

RESULT 6
 AAB15226
 ID AAB15226 standard; protein; 197 AA.
 XX
 AC AAB15226;
 XX
 DT 19-DEC-2000 (first entry)
 XX
 DE Hepatitis C virus NS4A-NS3 fusion protease #8.
 XX
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
 KW liver failure; liver cancer; mutant; mutein.
 XX
 OS Hepatitis C virus.
 OS Synthetic.
 XX
 PN WO2000040707-A1.
 XX
 PD 13-JUL-2000.
 XX
 PF 06-JAN-2000; 2000WO-US00345.
 XX
 PR 08-JAN-1999; 99US-0115271.
 XX
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX
 PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;

XX
 DR WPI: 2000-465976/40.
 DR N-PSDB; AAA73335.
 XX
 PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 PT
 XX
 PS Example 5; Fig 18; 66pp; English.
 XX
 CC The present sequence is a mutated version of a fusion protein created
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
 CC proteins are both essential for the replication of the virus, acting to
 CC cleave its replicative proteins from the polyprotein produced from the
 CC HCV genome. Inhibitors of the two proteins should be effective as
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A
 CC fusion proteins which can be used to identify inhibitors of this type, as
 CC well as enabling structural studies of the protease and
 CC protease:inhibitor complexes. This sequence contains the alpha-helix0
 CC wild-type sequence.
 XX
 SQ Sequence 197 AA;

Query Match 92.1%; Score 936; DB 21; Length 197;
 Best Local Similarity 93.9%; Pred. No. 5.4e-90;
 Matches 185; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKKGSVVIVGRINLSGDTAYAOQTREGCGTQKTSHTGRDNKNOVEGEVQIVSTATQTEFLA 60
 DB 1 MKKGSVVIVGRINLSGDTAYAOQTREGCGTQKTSHTGRDNKNOVEGEVQIVSTAAQTFLA 60
 QY 61 TSINGVLWTVYHGAGTRTASPKGPVTQMTYNDKDLVGMQAPQGSRSLSLTPCTCGSSDLY 120
 DB 61 TCINGVCTVYHGAGTRTASPKGPVTQMTYNDKDLVGMQAPQGSRSLSLTPCTCGSSDLY 120
 QY 121 LVTRHADVIPVRRGDSRGLSPRPISYLYKSGSGGPLLCPCPAGHAGVIFRAAVSTRGVAK 180
 DB 121 LVTRHADVIPVRRGDSRGLSPRPISYLYKSGSGGPLLCPCPAGHAGVIFRAAVSTRGVAK 180
 QY 181 AVDFIPVESLETTMRSP 197
 DB 181 AVDFIPVESLETTMRSP 197

RESULT 7
 AAB15220
 ID AAB15220 standard; protein; 195 AA.
 XX
 AC AAB15220;
 XX
 DT 19-DEC-2000 (first entry)
 XX
 DE Hepatitis C virus NS4A-NS3 fusion protease #2.
 XX
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
 KW liver failure; liver cancer; mutant; mutein.
 XX
 OS Hepatitis C virus.
 OS Synthetic.
 XX
 PN WO2000040707-A1.
 XX
 PD 13-JUL-2000.
 XX
 PF 06-JAN-2000; 2000WO-US00345.
 XX
 PR 08-JAN-1999; 99US-0115271.
 XX
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX

PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
 XX WPI: 2000-465976/40.
 DR N-PSDB: AAY73329.

XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 PT -
 XX
 PS Claim 23; Fig 12; 66pp; English.
 XX
 CC The present sequence is a mutated version of a fusion protein created
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
 CC proteins are both essential for the replication of the virus, acting to
 CC cleave its replicative proteins from the polyprotein produced from the
 CC HCV genome. Inhibitors of the two proteins should be effective as
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A
 CC fusion proteins which can be used to identify inhibitors of this type, as
 CC well as enabling structural studies of the protease and
 CC protease-inhibitor complexes. This sequence contains the alpha-helix0-1
 CC variant.
 XX

SQ Sequence 195 AA;
 Query Match 91.4%; Score 929; DB 21; Length 195;
 Best Local Similarity 92.9%; Pred. No. 2 9e-89;
 Matches 183; Conservative 3; Mismatches 9; Indels 2; Gaps 1;
 QY 1 MKKKGSVVIVGRINLSGDTAYAAQOTRGEQGTQKTSHTGRDKNQVEGEVQIVSTATQTFLA 60
 DB 1 MKKKGSVVIVGRIVLNG--AYAAQOTRGEQGTQKTSHTGRDKNQVEGEVQIVSTATQTFLA 58
 QY 61 TSINGVLWTVYHGACRTTIAASPKGPVTQMTYNTVDKDLVGVWAPQGSRSLSLTPTCTGSSDLY 120
 DB 59 TCINGVCMVTVYHGACRTTIAASPKGPVTQMTYNTVDKDLVGVWAPQGSRSLSLTPTCTGSSDLY 118
 QY 121 LVTRHADVIPVRRRGRSGSLSPRPISYLGSGGPGLLCPAGHAVGIFRAAVSTRGVAK 180
 DB 119 LVTRHADVIPVRRRGRSGSLSPRPISYLGSGGPGLLCPAGHAVGIFRAAVSTRGVAK 178
 QY 181 AVDFIPVESLETTMRSP 197
 DB 179 AVDFIPVESLETTMRSP 195

RESULT 8
 AAB15212
 ID AAB15212 standard; protein: 195 AA.
 XX
 AC AAB15212;
 XX
 DT 19-DEC-2000 (first entry)
 XX
 DE Hepatitis C virus NS4A-NS3 fusion protease #1.
 XX
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
 KW liver failure; liver cancer.
 XX
 OS Hepatitis C virus.
 OS Synthetic.
 XX
 PN WO2000040707-A1.
 XX
 PD 13-JUL-2000.
 XX
 PF 06-JAN-2000; 2000WO-US00345.
 XX
 PR 08-JAN-1999; 99US-0115271.
 XX
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.

XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
 XX WPI: 2000-465976/40.
 DR N-PSDB: AAY73328.

XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 PT -
 XX
 PS Example 2; Fig 10; 66pp; English.
 XX
 CC The present sequence is a fusion protein created using the Hepatitis C
 CC virus (HCV) NS3 and NS4A protease enzymes. These proteins are both
 CC essential for the replication of the virus, acting to cleave its
 CC replicative proteins from the polyprotein produced from the HCV genome.
 CC Inhibitors of the two proteins should be effective as antiviral
 CC treatments of HCV infection. This is useful as HCV can lead to chronic
 CC liver disease such as cirrhosis, liver failure and liver cancer. The
 CC present invention concerns a number of NS3 mutants and NS3-NS4A fusion
 CC proteins which can be used to identify inhibitors of this type, as well
 CC as enabling structural studies of the protease and protease-inhibitor
 CC complexes.
 XX

SQ Sequence 195 AA;
 Query Match 88.8%; Score 902; DB 21; Length 195;
 Best Local Similarity 91.9%; Pred. No. 2e-86;
 Matches 181; Conservative 1; Mismatches 13; Indels 2; Gaps 1;
 QY 1 MKKKGSVVIVGRINLSGDTAYAAQOTRGEQGTQKTSHTGRDKNQVEGEVQIVSTATQTFLA 60
 DB 1 MKKKGSVVIVGRIVLNG--AYAAQOTRGLLCIITSLTGRDKNQVEGEVQIVSTAAQTFLA 58
 QY 61 TSINGVLWTVYHGACRTTIAASPKGPVTQMTYNTVDKDLVGVWAPQGSRSLSLTPTCTGSSDLY 120
 DB 59 TCINGVCMVTVYHGACRTTIAASPKGPVTQMTYNTVDKDLVGVWAPQGSRSLSLTPTCTGSSDLY 118
 QY 121 LVTRHADVIPVRRRGRSGSLSPRPISYLGSGGPGLLCPAGHAVGIFRAAVSTRGVAK 180
 DB 119 LVTRHADVIPVRRRGRSGSLSPRPISYLGSGGPGLLCPAGHAVGIFRAAVSTRGVAK 178
 QY 181 AVDFIPVESLETTMRSP 197
 DB 179 AVDFIPVESLETTMRSP 195

RESULT 9
 AAY24943
 ID AAY24943 standard; protein: 665 AA.
 XX
 AC AAY24943;
 XX
 DT 07-SEP-1999 (first entry)
 XX
 DE HCV NS4A-NS3 complex SEQ ID NO:14.
 XX
 KW HCV; hepatitis C virus; single chain recombinant complex; linker;
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
 KW hydrophobic domain; covalent complex; detection; inhibitor.
 XX
 OS Hepatitis C virus.
 OS Synthetic.
 XX
 PN WO9928482-A2.
 XX
 PD 10-JUN-1999.
 XX
 PF 24-NOV-1998; 98WO-US24528.
 XX
 PR 28-JUL-1998; 98US-0094331.
 PR 28-NOV-1997; 97US-0067315.

XX (SCHE) SCHERING CORP.
 PA Malcolm BA, Taremi SS, Weber PC, Yao N;
 PI WPI; 1999-385385/32.
 XX New hepatitis C virus covalent complexes
 DR Claim 6; Page 90-92; 21lpp; English.
 PT The present invention describes a covalent hepatitis C virus (HCV)
 XX NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
 CC to the amino terminus of the HCV NS3 protease domain. The present
 CC sequence represents a specifically claimed example of the above
 CC complex. The covalent NS4A-NS3 complexes are useful for structural
 CC determination and determination of mode of binding of HCV inhibitors by
 CC NMR spectroscopy. They can also be used for detecting inhibitors of the
 CC protease activity, the helicase activity and the ATPase activity of NS3.
 CC The covalent NS4A-NS3 complexes are more soluble, stable and active than
 CC the non-covalent protease-peptide complexes previously available.
 XX
 SQ Sequence 665 AA;

Query Match 86.2%; Score 875.5; DB 20; Length 665;
 Best Local Similarity 85.2%; Pred. No. 6.8e-83;
 Matches 167; Conservative 15; Mismatches 11; Indels 3; Gaps 1;

QY 5 GSVVIVGRINLSGD---TAAQOTRGEQGTOKTSHTGRDKNQVEGEVQIVSTATQTFLAT 61
 DB 22 GSVVIVGRILSLSGSGSITAYSOOTRGLGCKKTSITGRDKNQVEGEVQIVSTATQTFLAT 81
 QY 62 SINGVLMTVYHGAGTRTIAISPKGPVTOMYTNVDKDLVGWQAPGQSRSLTPTCTCGSSDLYL 121
 DB 82 CVNGVCVTYHAGSKTLAGPKGPITOMYTNVDQDLVGWQAPPGARSILPTCTCGSSDLYL 141
 QY 122 VTRHADVIPVRRRGRSGSLSPRISYLGKSSGGPLLCPCAGHAGVIFRAAVSTRGVAKA 181
 DB 142 VTRHADVIPVRRRGRSGSLSPRISYLGKSSGGPLLCPCAGHAGVIFRAAVSTRGVAKA 201
 QY 182 VDFIPVESLETTMRSP 197
 DB 202 VDFVPVESMETTMRSP 217

RESULT 10
 AAY24947
 ID AAY24947 standard; Protein; 665 AA.
 XX
 AC AAY24947;
 XX
 DT 07-SEP-1999 (first entry)
 DE HCV NS4A-NS3 complex SEQ ID NO:18.
 XX
 KW HCV; hepatitis C virus; single chain recombinant complex; linker;
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
 KW hydrophobic domain; covalent complex; detection; inhibitor.
 XX
 OS Hepatitis C virus.
 OS Synthetic.
 XX
 PN WO9928482-A2.
 XX
 PD 10-JUN-1999.
 XX
 PF 24-NOV-1998; 98WO-US24528.
 XX
 PR 28-JUL-1998; 98US-0094331.
 PR 28-NOV-1997; 97US-0067315.
 XX

PA (SCHE) SCHERING CORP.
 XX Malcolm BA, Taremi SS, Weber PC, Yao N;
 PI WPI; 1999-385385/32.
 XX New hepatitis C virus covalent complexes
 DR Claim 6; Page 100-102; 21lpp; English.
 PT The present invention describes a covalent hepatitis C virus (HCV)
 XX NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
 CC to the amino terminus of the HCV NS3 protease domain. The present
 CC sequence represents a specifically claimed example of the above
 CC complex. The covalent NS4A-NS3 complexes are useful for structural
 CC determination and determination of mode of binding of HCV inhibitors by
 CC NMR spectroscopy. They can also be used for detecting inhibitors of the
 CC protease activity, the helicase activity and the ATPase activity of NS3.
 CC The covalent NS4A-NS3 complexes are more soluble, stable and active than
 CC the non-covalent protease-peptide complexes previously available.
 XX
 SQ Sequence 665 AA;

Query Match 85.9%; Score 872.5; DB 20; Length 665;
 Best Local Similarity 84.7%; Pred. No. 1.4e-82;
 Matches 166; Conservative 16; Mismatches 11; Indels 3; Gaps 1;

QY 5 GSVVIVGRINLSGD---TAAQOTRGEQGTOKTSHTGRDKNQVEGEVQIVSTATQTFLAT 61
 DB 22 GSVVIVGRILSLSGSGSITAYSOOTRGLGCKKTSITGRDKNQVEGEVQIVSTATQTFLAT 81
 QY 62 SINGVLMTVYHGAGTRTIAISPKGPVTOMYTNVDKDLVGWQAPGQSRSLTPTCTCGSSDLYL 121
 DB 82 CVNGVCVTYHAGSKTLAGPKGPITOMYTNVDQDLVGWQAPPGARSILPTCTCGSSDLYL 141
 QY 122 VTRHADVIPVRRRGRSGSLSPRISYLGKSSGGPLLCPCAGHAGVIFRAAVSTRGVAKA 181
 DB 142 VTRHADVIPVRRRGRSGSLSPRISYLGKSSGGPLLCPCAGHAGVIFRAAVSTRGVAKA 201
 QY 182 VDFIPVESLETTMRSP 197
 DB 202 VDFVPVESMETTMRSP 217

RESULT 11
 AAY24942
 ID AAY24942 standard; Protein; 665 AA.
 XX
 AC AAY24942;
 XX
 DT 07-SEP-1999 (first entry)
 DE HCV NS4A-NS3 complex SEQ ID NO:13.
 XX
 KW HCV; hepatitis C virus; single chain recombinant complex; linker;
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
 KW hydrophobic domain; covalent complex; detection; inhibitor.
 XX
 OS Hepatitis C virus.
 OS Synthetic.
 XX
 PN WO9928482-A2.
 XX
 PD 10-JUN-1999.
 XX
 PF 24-NOV-1998; 98WO-US24528.
 XX
 PR 28-JUL-1998; 98US-0094331.
 PR 28-NOV-1997; 97US-0067315.
 XX
 PA (SCHE) SCHERING CORP.

```
XX PI Malcolm BA, Taremi SS, Weber PC, Yao N;
XX DR WPI; 1999-385385/32.
XX PT New hepatitis C virus covalent complexes
XX PS Claim 6; Page 88-90; 21lpp; English.
XX CC The present invention describes a covalent hepatitis C virus (HCV)
XX CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
XX CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
XX CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
XX CC to the amino terminus of the HCV NS3 protease domain. The present
XX CC sequence represents a specifically claimed example of the above
XX CC complex. The covalent NS4A-NS3 complexes are useful for structural
XX CC determination and determination of mode of binding of HCV inhibitors by
XX CC NMR spectroscopy. They can also be used for detecting inhibitors of the
XX CC protease activity, the helicase activity and the ATPase activity of NS3.
XX CC The covalent NS4A-NS3 complexes are more soluble, stable and active than
XX CC the non-covalent protease-peptide complexes previously available.
XX SQ Sequence 665 AA;
Query Match 85.8%; Score 871.5; DB 20; Length 665;
Best Local Similarity 85.2%; Pred. No. 1.8e-82;
Matches 167; Conservative 14; Mismatches 12; Indels 3; Gaps 1;
Qy 5 GSVIVIGRINLSGD---TAYAQOTRGEQGTOKTSHTRDKNQVEGEVQIVSTATOTFLAT 61
Db ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Qy 62 SINGVLWTVYHGAGTRTITASPKGPVTOMYTNVDKDLVGWQAPQGSRSLTPTCTGSSDLYL 121
Db :||| ||||| :||| ||||| :||| ||||| :||| ||||| :||| ||||| :|||
Qy 82 CVNGVCWTVYHGAGSKTLAGPKGPITOMYTNVDQDLVGWQAPPGARSLSLTPTCTGSSDLYL 141
Qy 122 VTRHADVIPVRRRGDSRGSLLSPRPISYLGSGGGLLCPAGHAGVIFRAAVSTRGVAKA 181
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 182 VDFIPVESLETTMRSP 197
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 202 VDFVPVESMETTMRSP 217
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
RESULT 12
AA1780
ID AAY17880 standard; Protein; 216 AA.
XX AC AAY17880;
XX DT 07-SEP-1999 (first entry)
XX DE HCV NS4A-NS3 complex SEQ ID NO:4.
XX KW HCV; hepatitis C virus; single chain recombinant complex; linker;
XX KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
XX KW hydrophobic domain; covalent complex; detection; inhibitor.
XX OS Hepatitis C virus.
XX OS Synthetic.
XX PN WO9928482-A2.
XX PD 10-JUN-1999.
XX PF 24-NOV-1998; 98WO-US24528.
XX PR 28-JUL-1998; 98US-0094331.
XX PR 28-NOV-1997; 97US-0067315.
XX PA (SCHE ) SCHERING CORP.
XX PI Malcolm BA, Taremi SS, Weber PC, Yao N;
```

```
PI Malcolm BA, Taremi SS, Weber PC, Yao N;
XX DR WPI; 1999-385385/32.
XX PT New hepatitis C virus covalent complexes
XX PS Claim 6; Page 76-77; 21lpp; English.
XX CC The present invention describes a covalent hepatitis C virus (HCV)
XX CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
XX CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
XX CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
XX CC to the amino terminus of the HCV NS3 protease domain. The present
XX CC sequence represents a specifically claimed example of the above
XX CC complex. The covalent NS4A-NS3 complexes are useful for structural
XX CC determination and determination of mode of binding of HCV inhibitors by
XX CC NMR spectroscopy. They can also be used for detecting inhibitors of the
XX CC protease activity, the helicase activity and the ATPase activity of NS3.
XX CC The covalent NS4A-NS3 complexes are more soluble, stable and active than
XX CC the non-covalent protease-peptide complexes previously available.
XX SQ Sequence 216 AA;
Query Match 85.5%; Score 868.5; DB 20; Length 216;
Best Local Similarity 85.1%; Pred. No. 7.6e-83;
Matches 166; Conservative 15; Mismatches 11; Indels 3; Gaps 1;
Qy 5 GSVIVIGRINLSGD---TAYAQOTRGEQGTOKTSHTRDKNQVEGEVQIVSTATOTFLAT 61
Db ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Qy 62 SINGVLWTVYHGAGTRTITASPKGPVTOMYTNVDKDLVGWQAPQGSRSLSLTPTCTGSSDLYL 121
Db :||| ||||| :||| ||||| :||| ||||| :||| ||||| :||| ||||| :|||
Qy 82 CVNGVCWTVYHGAGSKTLAGPKGPITOMYTNVDQDLVGWQAPPGARSLSLTPTCTGSSDLYL 141
Qy 122 VTRHADVIPVRRRGDSRGSLLSPRPISYLGSGGGLLCPAGHAGVIFRAAVSTRGVAKA 181
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 182 VDFIPVESLETTMRS 196
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 202 VDFVPVESMETTMRSP 216
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
RESULT 13
AA1780
ID AAY24946 standard; Protein; 665 AA.
XX AC AAY24946;
XX DT 07-SEP-1999 (first entry)
XX DE HCV NS4A-NS3 complex SEQ ID NO:17.
XX KW HCV; hepatitis C virus; single chain recombinant complex; linker;
XX KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
XX KW hydrophobic domain; covalent complex; detection; inhibitor.
XX OS Hepatitis C virus.
XX OS Synthetic.
XX PN WO9928482-A2.
XX PD 10-JUN-1999.
XX PF 24-NOV-1998; 98WO-US24528.
XX PR 28-JUL-1998; 98US-0094331.
XX PR 28-NOV-1997; 97US-0067315.
XX PA (SCHE ) SCHERING CORP.
XX PI Malcolm BA, Taremi SS, Weber PC, Yao N;
```


[illegible]

Search completed: August 30, 2003, 19:12:25
Job time : 45.6227 secs

Db 1005 RRCEILLGPADGMVSKGWRLLAPITAYAAQOTRGLLGIITSLTGDKNOVEGEVQIVST 1064
QY 54 ATOTFLATSLNGVLTWYHAGTGTIASPKGPVTOMYTNVDKDLVGMWAPQGSRLTPTCT 113
Db 1065 AAOFLATCINGVCMVYHAGTGTIASPKGPVTOMYTNVDQDLVGMWAPQGSRLTPTCT 1124
QY 114 CGSSDLYLTRHADVTPVRRRGRSGLLSPRISYLGSSGGPILCPAGHAGVGFRAAV 173
Db 1125 CGSSDLYLTRHADVTPVRRRGRSGLLSPRISYLGSSGGPILCPAGHAGVGFRAAV 1184
QY 174 STRGVAKAVDFIPVESLETTMRSP 197
Db 1185 CTRGVAKAVDFIPVENLETTMRSP 1208

RESULT 2

S40770
genome polyprotein - hepatitis C virus
N:Contains: capsid protein C; envelope protein M; hepatitis virus (strain H)
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C:Accession: S40770; PC1285
R:Okamoto, H.
Submitted to the EMBL Data Library, March 1992
A:Reference number: S40770
A:Accession: S40770
A:Molecule type: genomic RNA
A:Residues: 1-3011 <OKA>
A:Cross-references: EMBL:D10749; NID:g221586; PID:BAA01582.1; PID:g221587
R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda, Jpn. J. Exp. Med. 60, 167-177, 1990
A:Title: The 5'-terminal sequence of the hepatitis C virus genome.
A:Reference number: PC1284; NID:91013116; PMID:2170712
A:Accession: PC1285
A:Molecule type: genomic RNA
A:Residues: 1-513 <OK2>
A:Cross-references: GB:D00831; NID:g221511; PIDN:BAA00705.1; PID:g221512
A:Experimental source: isolate HC-J1
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F:116-191/Product: capsid protein C #status predicted <CP>
F:192-389/Product: major envelope protein E #status predicted <EPM>
F:390-729/Product: major envelope protein E #status predicted <EPM>
F:730-1006/Product: nonstructural protein NS1 #status predicted <NS1>
F:1007-1615/Product: nonstructural protein NS2 #status predicted <NS2>
F:1230-1237/Product: hepatitis virus #status predicted <NS3>
F:1312-1317/Region: nucleotide-binding motif A (P-loop)
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS5>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 83.0%; Score 843.5; DB 1; Length 3011;
Best Local Similarity 82.4%; Pred. No. 5e-68;
Matches 168; Conservative 8; Mismatches 19; Indels 9; Gaps 1;

QY 3 KKGSVIVGRIN -----LSGDTAYAAQOTRGGTOKTSHTRGDKNOVEGEVQIVST 53
Db 1005 RKGEILLGPADGMVSKGWRLLAPITAYAAQOTRGLLGIITSLTGDKNOVEGEVQIVST 1064
QY 54 ATOTFLATSLNGVLTWYHAGTGTIASPKGPVTOMYTNVDKDLVGMWAPQGSRLTPTCT 113
Db 1065 AAOFLATCINGVCMVYHAGTGTIASPKGPVTOMYTNVDQDLVGMWAPQGSRLTPTCT 1124
QY 114 CGSSDLYLTRHADVTPVRRRGRSGLLSPRISYLGSSGGPILCPAGHAGVGFRAAV 173
Db 1125 CGSSDLYLTRHADVTPVRRRGRSGLLSPRISYLGSSGGPILCPAGHAGVGFRAAV 1184
QY 174 STRGVAKAVDFIPVESLETTMRSP 197
Db 1185 CTRGVAKAVDFIPVENLETTMRSP 1208

RESULT 3

GNWVCH
genome polyprotein - hepatitis C virus (strain H)
N:Contains: capsid protein C; envelope protein M; hepatitis virus (strain H)
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C:Accession: A36814; A41546
R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
submitted to Genbank, July 1992
A:Description: Genomic structure of the human prototype strain H of hepatitis C virus
A:Reference number: A36814
A:Accession: A36814
A:Molecule type: genomic RNA
A:Residues: 1-3011 <INC>
A:Cross-references: GB:M67463; NID:g329737; PIDN:AAA45534.1; PID:g329738
R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
A:Title: Genomic structure of the human prototype strain H of hepatitis C virus: comp
A:Reference number: A41546; NID:92052256; PMID:1658800
A:Contents: annotation
A:Note: neither amino acid nor nucleotide sequence is given
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruct
F:116-191/Product: capsid protein C #status predicted <CP>
F:192-389/Product: major envelope protein E #status predicted <EPM>
F:390-729/Product: major envelope protein E #status predicted <EPM>
F:730-1006/Product: nonstructural protein NS1 #status predicted <NS1>
F:1007-1615/Product: hepatitis virus #status predicted <NS2>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS5>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240

Query Match

Best Local Similarity 81.4%; Score 838.5; DB 1; Length 3011;
Matches 166; Conservative 10; Mismatches 19; Indels 9; Gaps 1;

QY 3 KKGSVIVGRIN -----LSGDTAYAAQOTRGGTOKTSHTRGDKNOVEGEVQIVST 53
Db 1005 RKGEILLGPADGMVSKGWRLLAPITAYAAQOTRGLLGIITSLTGDKNOVEGEVQIVST 1064
QY 54 ATOTFLATSLNGVLTWYHAGTGTIASPKGPVTOMYTNVDKDLVGMWAPQGSRLTPTCT 113
Db 1065 AAOFLATCINGVCMVYHAGTGTIASPKGPVTOMYTNVDQDLVGMWAPQGSRLTPTCT 1124
QY 114 CGSSDLYLTRHADVTPVRRRGRSGLLSPRISYLGSSGGPILCPAGHAGVGFRAAV 173
Db 1125 CGSSDLYLTRHADVTPVRRRGRSGLLSPRISYLGSSGGPILCPAGHAGVGFRAAV 1184
QY 174 STRGVAKAVDFIPVESLETTMRSP 197
Db 1185 CTRGVAKAVDFIPVENLETTMRSP 1208

RESULT 4

GNWVW
genome polyprotein - hepatitis C virus (strain Taiwan)
N:Contains: capsid protein C; envelope protein M; hepatitis virus (strain Taiwan)
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C:Accession: A40244
R:Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.
Virology 188, 102-113, 1992

A:Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the
A:Reference number: A40244; MUID:92230206; PMID:1314449
A:Accession: A40244
A:Molecule type: genomic RNA
A:Residues: 1-3010 <CHE>
A:Cross-references: GB:M4754
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
F:1-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polyprotein NS3 #status predicted <NS3>
F:1230-1237/Product: hepatitis C virus genome polyprotein NS4 #status predicted <NS4>
F:1312-1317/Product: hepatitis C virus genome polyprotein NS5 #status predicted <NS5>
F:1316-1319/Product: hepatitis C virus genome polyprotein NS6 #status predicted <NS6>
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,207

Query Match 81.3%; Score 826.5; DB 1; Length 3010;
Best Local Similarity 77.9%; Pred. No. 1.8e-66;
Matches 159; Conservative 18; Mismatches 16; Indels 9; Gaps 1;

QY 3 KKGSVVIVGRIN-----LSGDTAYAAQTRGEGTQKTSHTGRDKNOVEGEVQIVST 53
DB 1005 RRGREILLGPADSLGEGWRLAPITAYAAQTRGEGTQKTSHTGRDKNOVEGEVQIVST 1064
QY 54 ATOTFLATISINGVLWTVYHGAGTTRITASPKGPVTQMYTNVDKLVGWAQPGQSRSLTPCT 113
DB 1065 ATQSFLATCINGVCMVTFHGAGSKTLGPKGPITQMYTNVDQDLVGHAPPGARSLLTPCT 1124
QY 114 CGSSDLYLVTRHADVIPVRRGRSGSLSPRPISYLVKSGSGGGLPCPAGHAGVIFRAAV 173
DB 1125 CGSSDLYLVTRHADVIPVRRGRSGSLSPRPISYLVKSGSGGGLPCPAGHAGVIFRAAV 1184
QY 174 STRGVAKAVDFIPVESLETTMRSP 197
DB 1185 CTRGVAKAVDFIPVESMETTRSP 1208

RESULT 5
genome polyprotein - hepatitis C virus (strain JT)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein
C:Superfamily: hepatitis C virus genome polyprotein
F:1-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polyprotein NS3 #status predicted <NS3>
F:1230-1237/Product: hepatitis C virus genome polyprotein NS4 #status predicted <NS4>
F:1312-1317/Product: hepatitis C virus genome polyprotein NS5 #status predicted <NS5>
F:1316-1319/Product: hepatitis C virus genome polyprotein NS6 #status predicted <NS6>
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,207

Query Match 81.3%; Score 826.5; DB 1; Length 3010;
Best Local Similarity 77.9%; Pred. No. 1.8e-66;
Matches 159; Conservative 18; Mismatches 16; Indels 9; Gaps 1;

QY 3 KKGSVVIVGRIN-----LSGDTAYAAQTRGEGTQKTSHTGRDKNOVEGEVQIVST 53
DB 1005 RRGREILLGPADSLGEGWRLAPITAYAAQTRGEGTQKTSHTGRDKNOVEGEVQIVST 1064
QY 54 ATOTFLATISINGVLWTVYHGAGTTRITASPKGPVTQMYTNVDKLVGWAQPGQSRSLTPCT 113
DB 1065 ATQSFLATCINGVCMVTFHGAGSKTLGPKGPITQMYTNVDQDLVGHAPPGARSLLTPCT 1124
QY 114 CGSSDLYLVTRHADVIPVRRGRSGSLSPRPISYLVKSGSGGGLPCPAGHAGVIFRAAV 173
DB 1125 CGSSDLYLVTRHADVIPVRRGRSGSLSPRPISYLVKSGSGGGLPCPAGHAGVIFRAAV 1184
QY 174 STRGVAKAVDFIPVESLETTMRSP 197
DB 1185 CTRGVAKAVDFIPVESMETTRSP 1208

RESULT 5
genome polyprotein - hepatitis C virus (strain JT)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein
C:Superfamily: hepatitis C virus genome polyprotein
F:1-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polyprotein NS3 #status predicted <NS3>
F:1230-1237/Product: hepatitis C virus genome polyprotein NS4 #status predicted <NS4>
F:1312-1317/Product: hepatitis C virus genome polyprotein NS5 #status predicted <NS5>
F:1316-1319/Product: hepatitis C virus genome polyprotein NS6 #status predicted <NS6>
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,207

F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
Query Match 80.5%; Score 817.5; DB 1; Length 3010;
Best Local Similarity 76.5%; Pred. No. 1.2e-65;
Matches 156; Conservative 20; Mismatches 19; Indels 9; Gaps 1;

QY 3 KKGSVVIVGRIN-----LSGDTAYAAQTRGEGTQKTSHTGRDKNOVEGEVQIVST 53
DB 1005 RRGREILLGPADSLGEGWRLAPITAYAAQTRGEGTQKTSHTGRDKNOVEGEVQIVST 1064
QY 54 ATOTFLATISINGVLWTVYHGAGTTRITASPKGPVTQMYTNVDKLVGWAQPGQSRSLTPCT 113
DB 1065 ATQSFLATCINGVCMVTFHGAGSKTLGPKGPITQMYTNVDQDLVGHAPPGARSLLTPCT 1124
QY 114 CGSSDLYLVTRHADVIPVRRGRSGSLSPRPISYLVKSGSGGGLPCPAGHAGVIFRAAV 173
DB 1125 CGSSDLYLVTRHADVIPVRRGRSGSLSPRPISYLVKSGSGGGLPCPAGHAGVIFRAAV 1184
QY 174 STRGVAKAVDFIPVESLETTMRSP 197
DB 1185 CTRGVAKAVDFIPVESMETTRSP 1208

RESULT 6
genome polyprotein - hepatitis C virus
N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein
C:Superfamily: hepatitis C virus genome polyprotein
F:1-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polyprotein NS3 #status predicted <NS3>
F:1230-1237/Product: hepatitis C virus genome polyprotein NS4 #status predicted <NS4>
F:1312-1317/Product: hepatitis C virus genome polyprotein NS5 #status predicted <NS5>
F:1316-1319/Product: hepatitis C virus genome polyprotein NS6 #status predicted <NS6>
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,207

Query Match 80.1%; Score 813.5; DB 1; Length 3010;
Best Local Similarity 76.0%; Pred. No. 2.7e-65;
Matches 155; Conservative 21; Mismatches 19; Indels 9; Gaps 1;

QY 3 KKGSVVIVGRIN-----LSGDTAYAAQTRGEGTQKTSHTGRDKNOVEGEVQIVST 53
DB 1005 RRGREILLGPADSLGEGWRLAPITAYAAQTRGEGTQKTSHTGRDKNOVEGEVQIVST 1064
QY 54 ATOTFLATISINGVLWTVYHGAGTTRITASPKGPVTQMYTNVDKLVGWAQPGQSRSLTPCT 113
DB 1065 ATQSFLATCINGVCMVTFHGAGSKTLGPKGPITQMYTNVDQDLVGHAPPGARSLLTPCT 1124
QY 114 CGSSDLYLVTRHADVIPVRRGRSGSLSPRPISYLVKSGSGGGLPCPAGHAGVIFRAAV 173
DB 1125 CGSSDLYLVTRHADVIPVRRGRSGSLSPRPISYLVKSGSGGGLPCPAGHAGVIFRAAV 1184

A.Variety: isolate JX1

C>Date: 19-May-2000 Sequence_revision 19-May-2000 #text_change 23-Mar-2001

C:Accession: S18030; S35570; A48332; S18029

R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
submitted to the EMBL Data Library, September 1991

A:Description: A whole genome of hepatitis C virus cDNA was isolated from a single p

A:Reference number: S18028

A:Accession: S18030

A:Molecule type: genomic RNA

A:Residues: 1-3010 <HON>

A:Cross-references: EMBL:X61596; NID:g59478; PIDN:CAA43793.1; PID:g59479

A:Experimental source: isolate JX1 from an individual

R:Honda, M.; Kaneko, S.; Unoura, M.; Kobayashi, K.; Murakami, S.
Arch. Virol. 128, 163-169, 1993

A:Title: Sequence analysis of putative structural regions of hepatitis C virus isolate

A:Reference number: A48332; MUID:93119270; PMID:8380322

A:Accession: S33570

A:Molecule type: genomic RNA

A:Residues: 1-547, 'T', 549-621, 'V', 623-624, 'S', 626-652, 'DL', 655-761, 'T', 763-782 <HON>

A:Cross-references: EMBL:X61591

A:Note: this sequence is inconsistent with the nucleotide translation

A:Note: the authors translated the codon AGG for residue 43 as Pro, TGG for residue as Trp, and TTC for residue 771 as Ser

A:Note: sequence extracted from NCBI backbone (NCBIN:121747, NCBI:P:121748)

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; S

F:2-115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: envelope protein M #status predicted <EPM>

F:192-389/Product: major envelope protein E #status predicted <MEE>

F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F:1007-1615/Product: hepatitis C virus NS3 #status predicted <NS3>

F:1230-1237/Product: nucleotide-binding motif A (P-loop)

F:1312-1317/Region: nucleotide-binding motif B

F:1316-1319/Region: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>

F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate

Query Match	78.9%	Score	801.5	DB 1	Length	3010			
Best Local Similarity	75.5%	Pred. No.	3.3e-64						
Matches	154	Conservative	20	Mismatches	21	Indels	9	Gaps	1

Qy	3	KGSVVIVGRIN-----LSGDYTAQQTREGQGTQKTSHTGRDKNOVEGEIVST	53
Db	1005	KRGRELLGPDGDFEQGRWLLAPITAYSOOTRGLFCGCVTSUTGRDKNOVEGEIVST	1064
Qy	54	ATQTFATLSINGLVTVYHGAGTRTIAISPKGPVTOMTNYVDKDLVGHQAPSGSRSLTPCT	113
Db	1065	ATGSFLATCVNGCVTVYHGAGSKTLAGKPGINOMTNYVDODLVGHQAPSGAASLTPTCT	1124
Qy	114	CGSSDLYLVTRHADVIPVRRGDSRGLSPRPISYLGKSGGGPLLCAGHANGVIFRAAV	173
Db	1125	YGSSDLYLVTRHADVIPVRRGDSRGLSPRPVSYLKGSGGGPLCPGSHAVGIFRAAV	1184
Qy	174	STRGVAKAVDFIPVESLETTMRSP	197
Db	1185	CTRGVAKAVDFIPVESMETTMRSP	1208

RESULT 9

JC5620

genome polyprotein - hepatitis C virus (isolate EUH1480)

N:Contains: capsid protein C; envelope protein M; hepatitis C virus genotype 5a, the predominant

protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

C>Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001

C:Accession: JC5620

R:Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.
Biochem. Biophys. Res. Commun. 236, 44-49, 1997

A:Title: The complete coding sequence of hepatitis C virus genotype 5a, the predominant

A:Reference number: JC5620; MUID:97366593; PMID:9223423

F: 2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>
F: 196,209,234,305,325,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,20,22

Query Match 66.8% Score 679; DB 1: Length 3033;
Best Local Similarity 69.3%; Pred. No. 4,7e-53;
Matches 124; Conservative 27; Mismatches 28; Indels 0; Gaps 0;

QY 19 TAYAQOTRECGTQTSHTRDKNQVEGEVQIVSTATQTFIATSLNGVLWTVYHGAGTRPT 78
Db | ||||| : | : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
1034 TAYAQOTRGLLGTIVVMTGRDKEQAGEIQVLSLTGTTISGVLTWTVYHGAGNKT 1093
QY : ||||| : : ||||| : : ||||| : : ||||| : : ||||| : : ||||| : : ||||| :
179 IASPKGPVTQMTNVNDKDLVGQAPOGSRSLTPCTCGSGDLYLVTRHADVIPRRRGDSR 138
Db : ||||| : : ||||| : : ||||| : : ||||| : : ||||| : : ||||| : : ||||| :
1094 LAGSRGPVTQMTSSAEGDLLVGPSPCTCKLEPCTCGADVLYLVTRNADVIPARRGDKR 1153
QY 139 GSLLSPRTSYLKGGSGGPLLPACGHAVGIIPRAAVSTRGVAKAVDFIPVESLETMRSP 197
Db | ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
1154 GALLSPRLSTLKGSGGVPVLCPRGHAVGVFAAAVCSRGVAKSIDFIPVETLDIVTRSP 1212

RESULT 11
GNMVJ8
genome polyprotein - hepatitis C virus (strain HC-J8)
N:Contains: capsid protein C; envelope protein M; hepatitis virin (EC 3.4.21.98) (nonstr
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 31-Dec-1992 #sequence revision 31-Dec-1992 #text_change 19-Jan-2001
C:Accession: A40250; PQ0397; PQ0559
R:Okamoto, H.; Kural, K.; Okada, S.I.; Yamamoto, K.; Liuzuka, H.; Tanaka, T.; Fukuda
Virology 188, 331-341, 1992
A:title: Full-length sequence of a hepatitis C virus genome having poor homology to
A:Reference number: A40250; UID:92230232; PMID:1314459
A:Accession: A40250
A:Molecule type: genomic RNA
A:Residues: 1-3033 <OKA>
A:Cross-references: GB:D10988; GB:D01221; NID:g221608; PIDN:BAA01761.1; PID:g221609
R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap
J Gen Virol 73, 1131-1141, 1992
A:title: Analysis of a new hepatitis C virus type and its phylogenetic relationship.
A:Reference number: PQ0393; UID:92268871; PMID:1316939
A:Accession: PQ0397
A:Molecule type: genomic RNA
A:Residues: 2678-2754 <CHA>
A:Cross-references: DDBJ:D10134
A:Experimental source: isolate E-bl2
R:Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Mori, S.; Hijikata, M.; Shimom
Biochem Biophys Res Commun. 181, 279-285, 1991
A:title: Distribution of plural HCV types in Japan.
A:Reference number: PQ0554; UID:92068204; PMID:1720309
A:Accession: PQ0559
A:Molecule type: mRNA
A:Residues: 2678-2729 <KAT>
A:Cross-references: GB:D10562; GB:D90518; NID:g221523; PIDN:BAA01418.1; PID:g221524
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstru
F:1-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEP>
F:390-733/Product: nonstructural protein NS1 #status predicted <NS1>
F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>
F:1011-1619/Product: hepatitis virin #status predicted <NS3>
F:1234-1241/Region: nucleotide-binding motif A (P-loop)
F:1316-1321/Region: nucleotide-binding motif B
F:1320-1323/Region: DEXH motif
F:1620-1866/Product: nonstructural protein NS4a #status predicted <NA4>
F:1867-2017/Product: nonstructural protein NS4b #status predicted <NB4>
F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>
F: 196,209,233,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,20,22

A: Molecule type: DNA
A: Residues: 1-590 <PAR>
A: Cross-references: GB:AL162755; GB:AL157959; NID:g7379742; PIDN:CAB84658.1; PID:g738007
A: Experimental source: serogroup A, strain 22491
C: Genetics:
A: Gene: NMA1418
C: Superfamily: nitrate/nitrite sensor protein narX
C: Keywords: autophosphorylation; phosphohistidine; phosphoprotein; phosphotransferase; s
F:395/Active site: His (phosphohistidine intermediate) #status predicted

	Query Match	8.8%	Score 89.5;	DB 2;	Length 590;
	Best Local Similarity	20.8%	Pred. No. 2.9;		
	Matches 45;	Conservative 27;	Mismatches 79;	Indels 65;	Gaps 6;
QY	28	EQGTOKTSHTGRDKNQVEGVQIVSTATQTPLATSIINGVLWTVYHGAGTRTIASPKGPVT	87		
Db	213	EGGTPEFKQVGRCFNMGGRKILYDDLEGOVAEQ-----TRSLKQNONLT	259		
QY	88	QMYTNVDKDLVGWQAPQ-----GSRSLPTCTCGSSDLYLVTRHAD-----	127		
Db	260	LLY-QTTRDLHQSYIPQQAAEHFLNRLPAGADSGRVCLDGGSDVTVSIHHADCGTAAS	318		
QY	128	-----VIPVRRGDSRGSLSPRPTISYLKSGSGGPLICPAGHAVGIFRAAVSTR---	176		
Db	319	DLGKYHEEIFFIETQNETLGRLLLSFPNGISLDEDDRIILLQTLGRQLGVSLAGAKQEEK	378		
QY	177	-----GVAKAYDF--IPVESLET	192		
Db	379	RLLAVLQERNLIAQGLHDSIAQALTFLNLQVOMLET	414		

Search completed: August 30, 2003, 19:20:32
Job time : 17.2134 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 18:01:52 ; Search time 9.75674 Seconds
(without alignments)
949.524 Million cell updates/sec

Title: US-09-965-594-22

Perfect score: 1016

Sequence: 1 MKKGSVVIVGRINLSGDTA.....YAKAVDIPVESLETTMRSP 197

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	844.5	83.1	3011	1 POLG_HCV1	P26664 h genome po
2	838.5	82.5	3011	1 POLG_HCVH	P27958 h genome po
3	826.5	81.3	3010	1 POLG_HCVTW	P29846 h genome po
4	817.5	80.5	3010	1 POLG_HCVJT	Q00269 h genome po
5	813.5	80.1	3010	1 POLG_HCVBK	P26663 h genome po
6	813.5	80.1	3010	1 POLG_HCVJA	P26662 h genome po
7	679	66.8	3033	1 POLG_HCVJ6	P26660 h genome po
8	675	66.4	3033	1 POLG_HCVJ8	P26661 h genome po
9	90	8.9	321	1 HHQA_ARATH	Q9sel7 arabidopsis
10	85.5	8.4	209	1 PAAD_PSEAE	O9bx08 pseudomonas
11	83.5	8.2	437	1 DEGL_ARATH	Q22609 arabidopsis
12	83	8.2	452	1 AAMP_HUMAN	Q13685 homo sapien
13	78.5	7.7	485	1 Y136_TREPA	O83172 treponema p
14	78.5	7.7	660	1 VST2_HEVBU	P29326 hepatitis e
15	78.5	7.7	660	1 VST2_HEVPA	P33426 hepatitis e
16	78	7.7	401	1 FXH1_MOUSE	O88621 mus musculu
17	77.5	7.6	263	1 GRAK_MOUSE	O35205 mus musculu
18	77.5	7.6	301	1 MCP_BPF41	P25596 lactococcus
19	77.5	7.6	452	1 MLTD_ECOLI	P23931 escherichia
20	76.5	7.5	323	1 VPRD_SMRHV	P21407 squirrel mo
21	76.5	7.5	333	1 MOSA_RHIME	Q07607 rhizobium m
22	76	7.5	300	1 ERA_MYCLE	Q49768 mycobacteri
23	76	7.5	911	1 TB11_NEIMB	Q09056 neisseria m
24	76	7.5	3411	1 POLG_YEFV1	P03314 y genome po
25	76	7.5	3411	1 POLG_YEFV2	P19901 y genome po
26	76	7.5	3414	1 POLG_TBEVW	P14336 t genome po
27	75.5	7.4	248	1 TRY1_CHICK	Q90627 gallus gall
28	75.5	7.4	1425	1 NP4A_MOUSE	P59240 mus musculu
29	75.5	7.4	2269	1 WDR9_HUMAN	Q9nsi6 homo sapien
30	75	7.4	264	1 CTRL_HUMAN	P40313 homo sapien
31	75	7.4	300	1 SIAL_PIG	P31936 sus scrofa
32	75	7.4	467	1 NX1B_BOVIN	Q28142 bos taurus
33	74.5	7.3	248	1 GRAD_MOUSE	P11033 mus musculu

34	74.5	7.3	1155	1 POL_GALV	P21414 gibbon ape
35	74	7.3	659	1 VST2_HEVME	Q03500 hepatitis e
36	74	7.3	973	1 VP18_HUMAN	Q9P253 homo sapien
37	74	7.3	3414	1 POLG_TBEVH	Q01299 t genome po
38	73.5	7.2	248	1 TRY2_CHICK	Q90628 gallus gall
39	73.5	7.2	294	1 DPM1_USTMA	P54856 ustilago ma
40	73.5	7.2	352	1 SUB1_SINY3	Q01903 synechocyst
41	73.5	7.2	390	1 CS12_WHEAT	P46525 triticum ae
42	73.5	7.2	443	1 FLII_AQUAE	O67531 aquifex ae
43	73.5	7.2	485	1 VST2_HEVRH	Q00270 hepatitis e
44	73.5	7.2	660	1 VST2_HEVMY	Q04611 hepatitis e
45	73	7.2	478	1 MM03_RABIT	P28863 oryctolagus

ALIGNMENTS

RESULT 1

ID	POLG_HCV1	STANDARD;	PRT;	3011 AA.
AC	P26664;			
DT	01-AUG-1992 (Rel. 23, Created)			
DT	01-AUG-1992 (Rel. 23, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].			
DE	Hepatitis C virus (isolate 1) (HCV).			
OS	Hepatitis C virus (isolate 1) (HCV).			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.			
OX	NCBI_TaxID=11104;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	MEDLINE-01172826; PubMed-1848704;			
RA	Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C., Gallegos C., Colt D., Medina-Selby A., Barr P.J., Weiner A.J., Bradley D.W., Kuo G., Houghton M.;			
RT	*Genetic organization and diversity of the hepatitis C virus.*;			
CC	-1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION. NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.			
CC	-1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.			
CC	-1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate + (RNA)(N).			
CC	-1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND RNA.			
CC	-1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).			
CC	EMBL; M62321; AAA45676.1; --			
DR	PIR; A39166; GNVVC3.			
DR	PDB; 1ALV; 16-FEB-99.			
DR	PDB; 1HEI; 25-NOV-98.			
DR	MEROPS; S29.001; --			
DR	MEROPS; U39.001; --			
DR	InterPro; IPR001410; DEAD.			
DR	InterPro; IPR002522; HCV_capsid.			

DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_NS5b.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSvir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; Helicase_C; 1.
 DR Pfam: PF00998; Viral_RDRP; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXdc; 1.
 KW Polyprotein: Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein: Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane: Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 192 381 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 1007 1615 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 2014 3011 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 3011 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT TRANSMEM 347 369 POTENTIAL.
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 476 476 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRG64;
 Query Match 83.1%; Score 844.5; DB 1; Length 3011;
 Best Local Similarity 82.4%; Pred. No. 3,4e-70;
 Matches 168; Conservative 9; Mismatches 18; Indels 9; Gaps 1;
 3 KKGSVVIVGRIN-----LSGDYAYAOQTRGCGTQKTSHTGRKNQVEGEVQIVST 53

Db 1005 RRGREILLGPADGVNKGWLLAPITAYAAQTRGCGTQKTSHTGRKNQVEGEVQIVST 1064
 Qy 54 ATQTFLATISNGVLTYYHGAGTRTITASPKGPTQMYTNVDKDLGVQWQAPQSGRSLSPTCT 113
 Db 1065 AAQTFLATCINGVCMVYHGAGTRTITASPKGPTQMYTNVDKDLGVQWQAPQSGRSLSPTCT 1124
 Qy 114 CGSSDLYLVTRHADVIPVRRGDSRGLSPRISYLVKLGSSGGLLCPAGHAGVGFRAAV 173
 Db 1125 CGSSDLYLVTRHADVIPVRRGDSRGLSPRISYLVKLGSSGGLLCPAGHAGVGFRAAV 1184
 Qy 174 STRGVAKAVDFIPVESLETTMRSP 197
 Db 1185 CTGCVAKAVDFIPVENLETTMRSP 1208
 RESULT 2
 POLG_HCVH STANDARD; PRT; 3011 AA.
 ID AC P27958;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate H) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11108;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92052256; PubMed=1658800;
 RA Inchausti G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
 RA Prince A.M.;
 RT "Genomic structure of the human prototype strain H of hepatitis C
 RT virus: comparison with American and Japanese isolates.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
 RX MEDLINE=97331322; PubMed=9187654;
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
 RT "Structure of the hepatitis C virus RNA helicase domain.";
 RL Nat. Struct. Biol. 4:463-467(1997).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
 RX MEDLINE=98154321; PubMed=9493270;
 RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
 RA Murcko M.A., Lin C., Caron P.R.;
 RT "Hepatitis C virus NS3 RNA helicase domain with a bound
 RT oligonucleotide: the crystal structure provides insights into the mode
 RT of unwinding.";
 RL Structure 6:89-100(1998).
 CC [1-] FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
 CC [1-] FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
 CC [1-] FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
 CC ACTIVATION OF NS3.
 CC [1-] FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
 CC [1-] FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
 CC [1-] CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC [1-] CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC (RNA)(N).
 CC [1-] SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.

CC -1- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
 CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
 CC -1- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
 CC -1- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M67463; AAA45534.1; .
 CC PIR: A36814; GNMVCH.
 CC PDB: 1HEI; 25-NOV-98.
 CC PDB: 1A1V; 16-FEB-99.
 CC PDB: 1A1R; 17-JUN-98.
 CC MEROPS: S29.001; .
 CC MEROPS: U39.001; .
 CC TRANSFAC: T04155; .
 CC InterPro: IPR001410; DEAD.
 CC InterPro: IPR002522; HCV_capsid.
 CC InterPro: IPR002521; HCV_core.
 CC InterPro: IPR002519; HCV_env.
 CC InterPro: IPR002531; HCV_NS1.
 CC InterPro: IPR002518; HCV_NS2.
 CC InterPro: IPR004109; HCV_NS3.
 CC InterPro: IPR000745; HCV_NS4a.
 CC InterPro: IPR001490; HCV_NS4b.
 CC InterPro: IPR002868; HCV_NS5a.
 CC InterPro: IPR002166; HCV_RdRP.
 CC InterPro: IPR001650; Helicase_C.
 CC InterPro: IPR007095; RNA_pol_DS_Ps.
 CC InterPro: IPR007094; RNA_pol_Psvir.
 CC Pfam: PF01543; HCV_capsid; 1.
 CC Pfam: PF01542; HCV_core; 1.
 CC Pfam: PF01539; HCV_env; 1.
 CC Pfam: PF01560; HCV_NS1; 1.
 CC Pfam: PF01538; HCV_NS2; 1.
 CC Pfam: PF02907; HCV_NS3; 1.
 CC Pfam: PF01006; HCV_NS4a; 1.
 CC Pfam: PF01001; HCV_NS4b; 1.
 CC Pfam: PF01506; HCV_NS5a; 1.
 CC Pfam: PF00271; Helicase_C; 1.
 CC Pfam: PF00998; Viral_RdRP; 1.
 CC ProDom: PD186062; HCV_NS1; 1.
 CC SMART: SM00487; DEXdc; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolyase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 1 191 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 192 383 CAPSID PROTEIN C.
 FT CHAIN 384 746 ENVELOPE GLYCOPROTEIN E1.
 FT CHAIN 747 809 ENVELOPE GLYCOPROTEIN E2.
 FT CHAIN 810 1026 PROTEIN P7.
 FT CHAIN 1027 1657 NONSTRUCTURAL PROTEIN NS2.
 FT CHAIN 1658 1711 PROTEASE/HELICASE NS3.
 FT CHAIN 1712 1972 NONSTRUCTURAL PROTEIN NS4A.
 FT CHAIN 1973 2420 NONSTRUCTURAL PROTEIN NS4B.
 FT CHAIN 2421 3011 NONSTRUCTURAL PROTEIN NS5A.
 FT CHAIN 347 369 POTENTIAL.
 FT ACT_SITE 1083 1083 TRANSMEM.
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CHAIN 191 191
 FT CHAIN 383 383
 FT CHAIN 746 746
 FT CHAIN 809 809
 FT CHAIN 1026 1026
 FT CHAIN 1657 1657
 FT CHAIN 1711 1711
 FT CHAIN 1972 1972
 FT CHAIN 2420 2420
 FT CHAIN 3011 3011
 FT CHAIN 369 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234

FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 476 476
 FT CARBOHYD 532 532
 FT CARBOHYD 540 540
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT STRAND 1224 1226
 FT TURN 1232 1233
 FT TURN 1236 1238
 FT HELIX 1239 1246
 FT TURN 1247 1248
 FT STRAND 1251 1255
 FT HELIX 1258 1271
 FT TURN 1272 1272
 FT STRAND 1277 1280
 FT TURN 1281 1282
 FT STRAND 1283 1285
 FT STRAND 1291 1295
 FT HELIX 1296 1301
 FT TURN 1302 1303
 FT STRAND 1312 1316
 FT TURN 1317 1319
 FT HELIX 1323 1335
 FT TURN 1336 1340
 FT STRAND 1343 1347
 FT TURN 1352 1353
 FT TURN 1360 1361
 FT STRAND 1362 1366
 FT STRAND 1368 1368
 FT STRAND 1373 1375
 FT TURN 1376 1377
 FT STRAND 1378 1380
 FT HELIX 1382 1385
 FT STRAND 1389 1393
 FT HELIX 1397 1409
 FT TURN 1410 1411
 FT STRAND 1414 1417
 FT TURN 1419 1420
 FT STRAND 1432 1436
 FT TURN 1438 1439
 FT STRAND 1450 1453
 FT STRAND 1456 1463
 FT STRAND 1471 1478
 FT STRAND 1480 1480
 FT HELIX 1481 1488
 FT TURN 1489 1490
 FT STRAND 1497 1501
 FT STRAND 1507 1507
 FT STRAND 1511 1511
 FT HELIX 1514 1527
 FT STRAND 1532 1544
 FT STRAND 1550 1550
 FT HELIX 1555 1564
 FT HELIX 1570 1578
 FT TURN 1579 1580
 FT HELIX 1584 1597
 FT TURN 1598 1598
 FT HELIX 1606 1611
 FT TURN 1614 1618
 FT STRAND 1622 1623
 FT STRAND 1627 1627
 FT STRAND 1635 1636
 FT HELIX 1640 1652
 FT SEQUENCE 3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;
 Query Match 82.5%; Score 838.5; DB 1; Length 3011;
 Best Local Similarity 81.4%; Pred. No. 1.2e-69;

OY 3 KGSVVIVGRIN-----LSGDTAYAAQOTRGEQGTQKTSHTGRDNKNOVEGEVQIVST 53
DB 1005 RRGREILGLPADLEGGRWLLAPITAYAAQOTRGLFGCIITSLGRDNKNOVEGEVQIVST 1064
OY 54 ATQFTLATISINGLVTVYHAGTRTIASPKGPVTOMYTNVDKLVGQAPQGSRLTPCT 113
DB 1065 ATQSLATCINGCVTVYHAGSKTLAGPKPIOTMYTNVDLVGHPAGQCARSLTPCT 1124
OY 114 CGSSDLYLVTRHADVIPVRRGRSGSLSPRISYLYKSGSGGGLLCPAGHAVGIFRAAV 173
DB 1125 CGSSDLYLVTRHADVIPVRRGRSGSLSPRISYLYKSGSGGGLLCPGSHVGVIFRAAV 1184
OY 174 STRGVAKAVDFIPVSELTWNSP 197
DB 1185 CTRGVAKAVDFVPVSEMETWNSP 1208

RESULT 4

POLG_HCVJT STANDARD: PRT: 3010 AA.
AC Q00269;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirus)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-JT) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31642;
RN [1]
RP SEQUENCE FROM N.A. PubMed-1318627;
RX MEDLINE-92295714; PubMed-1318627;
RA Tanaka T., Kato M., Nakagawa M., Ootsuyama Y., Cho M.J.,
RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
RT "Molecular cloning of hepatitis C virus genome from a single Japanese
RT carrier: sequence variation within the same individual and among
RT infected individuals.";
RL Virus Res. 23:39-53(1992).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
CC (RNA)(N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D11168; BAA01943.1; -.
DR PIR: A45573; A45573.
DR PDB: 1AIQ; 25-MAR-98.
DR PDB: 1JXP; 14-JAN-98.
DR MEROPS: S29_001; -.
DR MEROPS: U39_001; -.
DR InterPro: IPR001410; DEAD.

DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_NS5a.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RdRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN 1007 1615 NON-STRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT CHAIN 2014 3010 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT TRANSMEM 347 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT ACT_SITE 1083 1083 POTENTIAL.
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 3010 AA; 326573 MW; 94A1C77435D642BB CRC64;

Query Match 80.5%; Score 817.5; DB 1; Length 3010;
Best Local Similarity 76.5%; Pred. No. 1.le-67;
Matches 156; Conservative 20; Mismatches 19; Indels 9; Gaps 1;
OY 3 KGSVVIVGRIN-----LSGDTAYAAQOTRGEQGTQKTSHTGRDNKNOVEGEVQIVST 53
InterPro: IPR001410; DEAD.


```

Db 1005 RRREILLPADIEGCGRLLAPITAYAQORGLGCVITSLTRDKNOVGEYQVYST 1064
Qy 54 ATOTFLATSLNGVLTWYHAGAGTRTIASPKGVTQMTYNVDKDLVGMQAPQGSRLTPT 113
Db 1065 ATOSFLATCVNGVCTVFGAGSKTLGPKPITQMTYNVDQDLVGMHAPPCCARSLTPT 1124
Qy 114 CGSSDLYLTHRADVTPVRRGRDGRSLSPRISVLYKGGSGGPLLCPAGHAGVIFRAAV 173
Db 1125 CGSSDLYLTHRADVTPVRRGRDGRSLSPRISVLYKGGSGGPLLCPAGHAGVIFRAAV 1184
Qy 174 STRGVAKAVDFIPVESLETTMRSP 197
Db 1185 CTRGVAKAVDFIPVESLETTMRSP 1208

RESULT 5
POLG_HCVBK STANDARD; PRT: 3010 AA.
AC P26663;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein p7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (PC 2.7.7.48)].
OS Hepatitis C virus (isolate BK) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11105;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-91140698; PubMed=1847440;
RA Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
RA Onishi E., Andoh T., Yoshida I., Okayama H.;
RT "Structure and organization of the hepatitis C virus genome isolated
RT from human carriers.";
RL J. Virol. 65:1105-1113(1991).
RN [2]
RP SEQUENCE OF 1487-1500.
RX MEDLINE-96235224; PubMed=8647104;
RA Borowski P., Helland M., Oehlmann K., Becker B., Kornetevy L.;
RA Moomaw E.W., Adachi T., Hostomska Z.;
RT "Non-structural protein 3 of hepatitis C virus inhibits
RT phosphorylation mediated by cAMP-dependent protein kinase.";
RL Eur. J. Biochem. 237:611-618(1996).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
RX MEDLINE-97015088; PubMed=8861916;
RA Love R.A., Parge H.E., Wickersham J.A., Hostomska Z., Habuka N.,
RA Moomaw E.W., Adachi T., Hostomska Z.;
RT "The crystal structure of hepatitis C virus NS3 proteinase reveals a
RT trypsin-like fold and a structural zinc binding site.";
RL Cell 87:331-342(1996).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
RX MEDLINE-98227846; PubMed=9568891;
RA Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
RA Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;
RT "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C
RT virus: a 2.2-A resolution structure in a hexagonal crystal form.";
RL Protein Sci. 7:837-847(1998).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
CC [RNA](N).
```

```

CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M58335; AAA72945.1; -.
CC FIR: A38465; GNVYTC.
CC PDB: 1AIQ; 25-MAR-98.
CC PDB: 1JXP; 14-JAN-98.
CC PDB: 1NS3; 08-APR-98.
CC PDB: 1C2P; 15-NOV-00.
CC PDB: 1CSJ; 08-NOV-99.
CC PDB: 1GX5; 09-APR-02.
CC PDB: 1GX6; 10-APR-02.
CC PDB: 1QVU; 26-JUN-00.
CC PDB: 8OHM; 20-APR-99.
CC MEROPS: S29.001; -.
CC
CC InterPro: IPR001410; DEAD.
CC InterPro: IPR002522; HCV_capsid.
CC InterPro: IPR002521; HCV_core.
CC InterPro: IPR002519; HCV_env.
CC InterPro: IPR002531; HCV_NS1.
CC InterPro: IPR002518; HCV_NS2.
CC InterPro: IPR004109; HCV_NS3.
CC InterPro: IPR000745; HCV_NS4a.
CC InterPro: IPR001490; HCV_NS4b.
CC InterPro: IPR002868; HCV_NS5a.
CC InterPro: IPR002166; HCV_RORP.
CC InterPro: IPR007095; RNA_pol_DS_PS.
CC InterPro: IPR007094; RNA_pol_PSVir.
CC Pfam: PF01543; HCV_capsid; 1.
CC Pfam: PF01542; HCV_core; 1.
CC Pfam: PF01539; HCV_env; 1.
CC Pfam: PF01560; HCV_NS1; 1.
CC Pfam: PF01538; HCV_NS2; 1.
CC Pfam: PF02907; HCV_NS3; 1.
CC Pfam: PF01006; HCV_NS4a; 1.
CC Pfam: PF01001; HCV_NS4b; 1.
CC Pfam: PF01506; HCV_NS5a; 1.
CC Pfam: PF00998; Viral_RORP; 1.
CC ProDom: PD186062; HCV_NS1; 1.
CC SMART; SM00487; DEXDC; 1.
CC PolyProtein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
CC Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
CC Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
CC 3D-structure. 1 1
CC INIT_MET 1 1
CC CHAIN 1 115
CC CHAIN 116 191
CC CHAIN 192 383
CC CHAIN 384 729
CC CHAIN 730 1006
CC CHAIN 1007 1615
CC CHAIN 1616 1862
CC CHAIN 1863 2013
CC CHAIN 2014 3010
CC TRANSMEM 347 369
CC ACT_SITE 1083 1093
CC ACT_SITE 1107 1107
CC ACT_SITE 1165 1165
CC NP_BIND 1230 1237
CC SITE 1316 1319
```

FT CARBOHYD 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1031 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX 1039 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1050 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1059 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1068 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1075 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1077 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX 1082 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1086 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1090 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1093 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1095 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1101 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1104 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1108 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1120 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1122 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1129 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1135 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1139 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1149 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1158 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX 1162 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1168 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1172 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1175 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1187 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1189 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX 1198 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1203 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1680 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3010 AA; F8422D5ECCFDF9C CRC64;

Query Match 80.18; Score 813.5; DB 1; Length 3010;
Best Local Similarity 76.08; Pred. No. 2.7e-67;
Matches 155; Conservative 21; Mismatches 19; Indels 9; Gaps 1;
Qy 3 KKGWVIVGRIN-----LSGDTAYAQOTRGEOCTQKTSHTGRDKNOVEGEVQIVST 53
Db 1005 RRGKEILLGADSLGGRGLLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQIVST 1064
Qy 54 ATQTFATSLVINGVWTVYHGAGTRTITASPKGVPTQMTYNDKDLVGHQAPQOGSRSLTPCT 113
Db 1065 ATQTFATSLVINGVWTVYHGAGTRTITASPKGVPTQMTYNDKDLVGHQAPQOGSRSLTPCT 1124
Qy 114 CGSSDLYLVTRHADVIVPVRGDSRGSLLSPRPISYLGSSGGPGLLCGAGHANGVIFRAAV 173
Db 1125 CGSSDLYLVTRHADVIVPVRGDSRGSLLSPRPISYLGSSGGPGLLCGAGHANGVIFRAAV 1184
Qy 174 STRGVAKAVDFIPVESLETTMRSP 197
Db 1185 CTRGVAKAVDFIPVESLETTMRSP 1208

RESULT 6

ID POLG_HCVJA STANDARD; PRT; 3010 AA.
AC P26662;
DT 01-AUG-1992 (Rel. 23, Created)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate Japanese) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91088550; PubMed=2175903;
RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
RA Sugimura T., Shimotohno K.;
RT "Molecular cloning of the human hepatitis C virus genome from
RT Japanese patients with non-A, non-B hepatitis.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
RN [2]
RP DISCUSSION OF SEQUENCE.
RX MEDLINE=91192160; PubMed=1849488;
RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraishi K.,
RA Ohkoshi S., Shimotohno K.;
RL FEBS Lett. 280:325-328(1991).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
CC (RNA)(N).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D90208; BAAL4233.1; -
CC PIR; A39253; GNMVJ3.
CC HSP; P26663; LTXP.
CC MEROPS; S29.001; -
CC MEROPS; O39.001; -
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR002522; HCV_capsid.
CC InterPro; IPR002521; HCV_core.
CC InterPro; IPR002531; HCV_env.
CC InterPro; IPR002518; HCV_NS1.
CC InterPro; IPR004109; HCV_NS3.
CC InterPro; IPR000745; HCV_NS4a.
CC InterPro; IPR001490; HCV_NS4b.
CC InterPro; IPR002868; HCV_NS5a.
CC InterPro; IPR002166; HCV_RdRp.

InterPro: IPR001650; Helicase_C.
 InterPro: IPR007095; RNA_pol_DS_PS.
 InterPro: IPR007094; RNA_pol_PSVir.
 Pfam: PF01543; HCV_capsid; 1.
 Pfam: PF01542; HCV_core; 1.
 Pfam: PF01539; HCV_env; 1.
 Pfam: PF01560; HCV_NS1; 1.
 Pfam: PF01538; HCV_NS2; 1.
 Pfam: PF02907; HCV_NS3; 1.
 Pfam: PF01006; HCV_NS4a; 1.
 Pfam: PF01001; HCV_NS4b; 1.
 Pfam: PF01506; HCV_NS5a; 1.
 Pfam: PF00271; helicase_C; 1.
 Pfam: PF00998; Viral_RdRP; 1.
 Pfam: PF0186062; HCV_NS1; 1.
 SMART: SM00487; DEXDC; 1.
 Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT CHAIN 3010 369
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 532 532
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2788 2788
 FT CARBOHYD 3010 3010
 FT SEQUENCE 3010 AA; 327017 MW; AA993794F46DB185 CRC64;
 Query Match 80.18; Score 813.5; DB 1; Length 3010;
 BestLocal Similarity 75.08; Pred. No. 2.7e-67;
 Matches 153; Conservative 23; Mismatches 19; Indels 9; Gaps 1;
 Oy 3 KKGSWIVGRINLSGD-----TAYAOQTRGEOCTQKTSHTGRDKNOVEQVIST 53
 Db 1005 RRGREILLGPADSGEGQWRLLAPITAYSQOTRGLLCIITSITGRDKNOVDEVOVLST 1064
 Oy 54 ATQFLATSTINGVLWYTHGAGTRTITASPKGPTVMTYNDKLVGQAPQGSRSITPCT 113
 Db 1065 ATQSFLLATVNGVCTWYTHGAGSKTLAGPKGPTVMTYNDKLVGQAPQGSRSITPCT 1124
 Oy 114 CGSSDLVLTTHADVIPIVRRGDSRGLLSRPRTISYLKSGSGPLLCPAGHAGVIFRAAV 173
 Db 1125 CGSSDLVLTTHADVPIVRRGDSRGLLSRPRTISYLKSGSGPLLCPAGHAGVIFRAAV 1184
 Oy 174 STRGVAKAVDFIPVESLETTMRSP 197

Db 1185 CTGRVAKAVDFIPVESMETTMRSP 1208
 RESULT 7
 POLG_HCVJ6 STANDARD; PRT; 3033 AA.
 AC P26660;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.-); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-J6) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID-11113;
 RX MEDLINE-92044440; PubMed-1658196;
 RA Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Iizuka H.,
 RA Machida A., Miyakawa Y., Mayumi M.;
 RT "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
 RT from a human carrier: comparison with reported isolates for conserved
 RT and divergent regions.";
 RL J. Gen. Virol. 73:2697-2704(1991)
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 CC {RNA}(N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; D00944; BAA00792.1; -
 CC PIR; JQ1303; JQ1303.
 CC HSSP; P27958; 1REI.
 CC MEROPS; S29.001; -
 CC MEROPS; U39.001; -
 CC InterPro: IPR001410; DEAD.
 CC InterPro: IPR002522; HCV_capsid.
 CC InterPro: IPR002521; HCV core.
 CC InterPro: IPR002519; HCV env.
 CC InterPro: IPR002531; HCV NS1.
 CC InterPro: IPR002518; HCV NS2.
 CC InterPro: IPR004109; HCV NS3.
 CC InterPro: IPR001490; HCV NS4a.
 CC InterPro: IPR000745; HCV NS4b.
 CC InterPro: IPR002868; HCV NS5a.
 CC InterPro: IPR002166; HCV_RdRP.
 CC InterPro: IPR001650; Helicase_C.
 CC InterPro: IPR007095; RNA_pol_DS_PS.
 CC InterPro: IPR007094; RNA_pol_PSVir.
 CC Pfam; PF01543; HCV_capsid; 1.

ProDom: PD186062; HCV_NS1: 1.
DR SHART: SH00487; DEXdc: 1.
 KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT CHAIN 3034 3301
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1089
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 233 233
 FT CARBOHYD 299 299
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CARBOHYD 627 627
 FT CARBOHYD 649 649
 FT CARBOHYD 1091 1091
 FT CARBOHYD 2038 2038
 FT CARBOHYD 2359 2359
 FT CARBOHYD 2811 2811
 SQ SEQUENCE 3033 AA: 330177 MW: 1A173E7E3381FD1A CRC64;
 Query Match 66.48; Score 675; DB 1; Length 3033;
 Best Local Similarity 69.84; Pred. No. 2,1e-54;
 Matches 125; Conservative 24; Mismatches 30; Indels 0; Gaps 0;
 QY 19 TAYAQOTRGEGTOKTSHTGRDNQKQVEQIVSTATOTFLATSLVGVVHGAGT 78
 DB 1034 TAYTQOTRGLLGAIVSVLTGRDNQKQVGVLSVTVQIFLCTISGLVTVVHGAGNKT 1093
 QY 79 IASPKGPVTQYTNVNDKLVQWQAPQGSRLSTPTCGSSDLYLVTRHADVIPVRRGDSR 138
 DB 1094 LAGPKGPVTQYTNVNDKLVQWQAPQGSRLSTPTCGSSDLYLVTRHADVIPVRRGDSR 1153
 QY 139 GLLSPRISYILKSSGSPILCPAGHAGVIFRAAVSTRGVAKAVDFIPVESLETMRSP 197
 DB 1154 GALLSPRLSTLKGSSGSPVLCRSHAGVGLFRAAVCARGVAKSIDFIPVESLDVATRP 1212
 RESULT 9
 ID HHOAARATH STANDARD; PRT: 321 AA.
 AC Q9SEL7; O49507;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Protease HhoA, chloroplast precursor (EC 3.4.21.-).
 GN HHOA OR AT4G18370 OR F28J12.30.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II: Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Lensch M.H.A., Sokolenko A., Herrmann R.G.;
 RT "Identification and characterization of the chloroplast HhoA protease,
 a homolog to the bacterial periplasmic protease HhoA";
 FT Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 RL (2)
 RN SEQUENCE FROM N.A.
 RP STRAIN=cv. Columbia;
 RC MEDLINE=20083488; PubMed=10617198;
 RX Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
 RA Harris B., Ansoerger W., Brandt P., Grivell L., Rieger M.,
 RA Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
 RA Kreis B., Delseny M., Puigdomenech P., Watson M., Schmidtheini T.,
 RA Vos P., Hohelsel J., Zimmermann W., Wedler H., Ridley P.,
 RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
 RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
 RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
 RA Weitzenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
 RA Holzner E., Brandt A., Peters S., van Staveren M., Dirkse W.,
 RA Mooijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
 RA Berner S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
 RA De Keyser A., Buysshaert C., Gielen J., Villarroel R., De Clercq R.,
 RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
 RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,
 RA Pettitt A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
 RA Borkova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,
 RA Dose S., de Haan M., Maarle A., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,
 RA Neubann C., Argizou A., Vitale D., Liguori R., Piravandi E.,
 RA Massenot O., Quigley F., Clabaud G., Muendlein A., Feilber R.,
 RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
 RA Chefdor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
 RA Gibbons T., Weber N., Vandenbol M., Barques M., Terol J., Torres A.,
 RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
 RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Biele C.,
 RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
 RA Zaccaria L., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
 RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
 RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
 RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
 RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
 RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
 RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
 RA Nelson J., Splith J., Ryan E., Andrews S., Geisel C., Layman D.,
 RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
 RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Jordan C.,
 RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
 RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
 RA Chen E., Marra M., Martienssen R., McCombie W.R.;
 RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
 thaliana";
 RN Nature 402:769-777(1999).
 RP (3)
 SEQUENCE OF 72-82; 96-110; 150-159; 178-211 AND 306-320.
 RA Schubert M., Peterson U., Funk C., Haas B., Schroeder W.P.,
 RA Kieselbach T.;
 RT "The chloroplast lumen from Arabidopsis thaliana";
 CC 1- SUBCELLULAR LOCATION: Chloroplast; within the thylakoid lumen.
 CC 1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
 CC 1- CAUTION: Ref.2 sequences differ from that shown due to erroneous
 CC gene model prediction. AT4G18370 and AT4G18375 were originally
 CC fused into a single gene.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF114386; AAF24060.1; -
EMBL; AL021710; CAA16717.1; ALT_SEQ.
EMBL; AL161548; CAB78839.1; ALT_SEQ.
MEROPS; S01.279; -
InterPro; IPR001940; Protease2C.
InterPro; IPR001254; Ser_protease_Try.
Pfam; PF00089; trypsin; 1.
PRINTS; PR00834; PROTESAS2C.
KW Hydrolase; Serine protease; Chloroplast; Thylakoid; Transit peptide.
FT TRANSIT 1 26 CHLOROPLAST (POTENTIAL).
FT TRANSIT 27 71 THYLAKOID.
FT CHAIN 72 321 PROTEASE HHOA.
FT DOMAIN 77 87 POLY-GLU.
FT ACT_SITE 145 145 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 186 186 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 264 264 CHARGE RELAY SYSTEM (POTENTIAL).
FT CONFLICT 40 40 R -> G (IN REF 1).
SQ SEQUENCE 321 AA; 34691 MW; 68DB81E0BD27A7A7 CRC64;

Query Match 8.9%; Score 90; DB 1; Length 321;
Best Local Similarity 23.8%; Pred. No. 0.43;
Matches 55; Conservative 31; Mismatches 89; Indels 56; Gaps 11;

QY 2 KKGSVVIVGRINL-----SGDTAAQTRGSGT-----QKTSHTGRDKNOVEGEVQIV 51
DB 95 KTSVSVVYIEAIELPKTSGGDILTDENKIEGTGSGFVWDKLGHI-----VTNYHVIA 148
QY 52 STATOTFLATISNGVLTVYHGAGTRTIASPKGPVTQMYTNVDKDLVGWQAPQGSRLTP 111
DB 149 KLATDQF---GLORCKVSLVDKGR--FSKEGKIVGL--DPDNDLAVLKIEGRELNP 201
QY 112 CTCGSSDLVYTRHADVIPVRRRGDSR-----SLLSPRPISYLK----- 151
DB 202 VVLGTSNDRVGQSCFAI-----GNPYGENTLTIGVVGSLGRLPEIPSPNGKSISEAQTQ 256
QY 152 -----GSSGGPLCPAGHAGVIFRAAVSTR--GVAKAVDF-IPVESLETTM 194
DB 257 ADINSNGSGPLDSYGHITGVTATFTKSGMSSGVNFAIPDTVVRTV 307

RESULT 10
ID PAAD_PSEAE STANDARD; PRT; 209 AA.
AC Q9HX08.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Probable aromatic acid decarboxylase (EC 4.1.1.-).
GN PA4019.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-O.T., Erwin A.L., Miziochuch S.D., Warren P., Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M., Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y., Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M., Smith K.A., Spencer D.H., Wong K.K.-S., Wu Z., Paulsen I.T., Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -!- SIMILARITY: BELONGS TO THE POLYPRENYL P-HYDROXYBENZOATE / PHENYLACRYLIC ACID DECARBOXYLASES FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF004818; AAG07406.1; -
PIR; H83144; H83144.
InterPro; IPR003382; Flavoprotein.
Pfam; PF02441; Flavoprotein; 1.
KW Hypothetical protein; Lyase; Decarboxylase; Complete proteome.
SQ SEQUENCE 209 AA; 22367 MW; 01FD081CC495D3F6 CRC64;

Query Match 8.4%; Score 85.5; DB 1; Length 209;
Best Local Similarity 27.9%; Pred. No. 0.67;
Matches 51; Conservative 16; Mismatches 61; Indels 55; Gaps 11;

QY 43 QVEGEVQ-IVSTATOTFLATISNGVL-----WIVYHGAGTRTIASPKGPVTQMT 91
DB 29 QEREVHFLISKAAQLVMATETVALPAKPOAQAFLEYCGAAGQI-----RVFG 80
QY 92 NYDKDLVGWQAPQGSRLTP-----CTCGSSDL-----YLVTRHADVIPVRRRGDS 137
DB 81 QND-----WMAPASGSSAPNAMYICPSTGTLSAVATGACNNLIERAADVALKER---- 131
QY 138 RGSLLSPR--PIS-----YKSSGGPLCPAGHAGVIFRAAVSTRGVAKAVDFIPVES 189
DB 132 RPLVLPVREAPFSSIHLENKLSNLGAVILPA--APGFYH---QPQSVEDLVDFVARI 186
QY 190 LET 192
DB 187 LNT 189

RESULT 11
ID DEGLARATH STANDARD; PRT; 437 AA.
AC 022609; O9LK85;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protease De-like 1, chloroplast precursor (EC 3.4.21.-).
GN DEGP1 OR DEGP OR AT3G27925 OR K16N12.18.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A. AND CHARACTERIZATION.
RX MEDLINE=98175982; PubMed=9507020;
RA Itzhaki H., Naveh L., Lindahl M., Cook M., Adam Z.;
RT "Identification and characterization of Degp, a serine protease associated with the luminal side of the thylakoid membrane.";
RL J. Biol. Chem. 273:7094-7098(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20363099; PubMed=10907853;
RA Kaneko T., Katoh T., Sato S., Nakamura A., Asamizu E., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II. Sequence features of the 4,251,695 bp regions covered by 90 P1, TAC and BAC clones.";
RL DNA Res. 7:217-221(2000).
RN [3]
RP SEQUENCE OF 104-118.
RC STRAIN=cv. Columbia;
RA Kieselbach T., Bystedt M., Schroeder W.P.;
RL Submitted (JUL-2000) to the SWISS-PROT data bank.
CC -!- FUNCTION: SERINE PROTEASE THAT IS REQUIRED AT HIGH TEMPERATURE.

MAY BE INVOLVED IN THE DEGRADATION OF DAMAGED PROTEINS. IN VIVO.
CAN DEGRADE BETA-CASIN.
-1- ENZYME REGULATION: INHIBITED BY PHENYLMETHYLSULFONYL FLUORIDE AND O-PHENANTHROLINE.
-1- SUBCELLULAR LOCATION: BOUND TO LUMINAL SIDE OF THE THYLAKOID MEMBRANE.
-1- INDUCTION: By heat shock.
-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
-1- SIMILARITY: Contains 1 PDZ/DHR domain.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF028842; AAC39436.1; .
EMBL; AP000371; BAB02539.1; .
EMBL; AP001302; BAB02539.1; JOINED.
MEROPS: S01.279; .
InterPro: IPR001478; PDZ.
InterPro: IPR001940; Protease2C.
InterPro: IPR001254; Ser-protease_Try.
Pfam: PF00595; PDZ; 1.
Pfam: PF00089; trypsin; 1.
PRINTS: PR00834; PROTEASES2C.
SMART: SM00228; PDZ; 1.
PROSITE; PS0106; PDZ; 1.
KW Hydrolyase; Serine protease; Transit peptide; Chloroplast; Thylakoid.
FT TRANSIT 1 ? CHLOROPLAST (POTENTIAL).
FT TRANSIT 2 ? THYLAKOID.
FT CHAIN 104 437 PROTEASE DO-LIKE 1.
FT DOMAIN 152 321 SERINE PROTEASE.
FT DOMAIN 324 421 PDZ.
FT ACT_SITE 171 171 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 201 201 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 280 280 CHARGE RELAY SYSTEM (POTENTIAL).
FT CONFLICT 12 23 HSPSPQLSNST -> SSTFLHSPSSHL (IN REF. 2).
FT CONFLICT 36 36 V -> I (IN REF. 2).
FT CONFLICT 54 54 P -> S (IN REF. 2).
FT CONFLICT 60 60 G -> R (IN REF. 2).
FT CONFLICT 64 64 G -> D (IN REF. 2).
FT CONFLICT 68 69 LL -> HF (IN REF. 2).
FT CONFLICT 355 355 L -> V (IN REF. 2).
FT CONFLICT 381 381 I -> V (IN REF. 2).
FT CONFLICT 416 416 Q -> E (IN REF. 2).
SQ SEQUENCE 437 AA; 48213 MW; 1497B1AB3F5FF2A4 CRC64;

Query Match 8.2%; Score 83.5; DB 1; Length 437;
Best Local Similarity 26.2%; Pred. No. 2.5;
Matches 45; Conservative 17; Mismatches 55; Indels 55; Gaps 7;

QY 70 VYHGAGTRTITASPKGPVTQMY-----TNVDKDLVGV-----QA 102
DB 150 VPGSGGFVMDQGHVITVYHVRGASDLRVTLADQTTDFDAKVGVGDQDKVAVLRIDA 209
QY 103 PGGRSLTPTCTGSSDLYLV-----TRHADVIPVRRRGDSRGLSPRPI 147
DB 210 PK-NKLRPIPVGVSDLLVGVQKVFAGNPFGLDHLTLTGVISGLRRETS--SAATGRPI 265
QY 148 SYL-----KGSGGLPLCPAGHAVIPRAAVSTRGVAKAVDF-IPVESL 190
DB 266 QDVITDAAINPGNSGGPLDSSGTLIGINTAIYSPGASSGVGFSPIDTV 317

RESULT 12
AAMP_HUMAN STANDARD; PRT; 452 AA.
ID AAMP_HUMAN
AC Q13685;
DT 15-JUL-1998 (Rel. 36, Created)

15-JUL-1998 (Rel. 36, Last sequence update)
28-FEB-2003 (Rel. 41, Last annotation update)
DE Angio-associated migratory cell protein.
GN AAMP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Brain;
RX MEDLINE=95262124; PubMed=7743515;
RA Beckner M.E., Kruttsch H.C., Stracke M.L., Williams S.T.,
RA Gallardo J.A., Liotta L.A.;
RT Identification of a new immunoglobulin superfamily protein expressed
RT in blood vessels with a heparin-binding consensus sequence.*;
RL Cancer Res. 55:2140-2149(1995).
CC -!- FUNCTION: MAY HAVE A FUNCTION IN MIGRATING CELLS.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN BLOOD VESSELS. STRONGLY EXPRESSED
CC IN ENDOTHELIAL CELLS, CYTOTROPHOBLASTS, AND POORLY DIFFERENTIATED
CC COLON ADENOCARCINOMA CELLS FOUND IN LYMPHATICS.
CC -!- SIMILARITY: Contains 8 WD repeats.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M95627; AAA68889.1; .
PIR; I39383; I39383.
GeneW; HGNC:18; AAMP.
MIM; 603488; .
GO; GO:0008201; F:heparin binding activity; TAS.
InterPro: IPR001880; WD40.
Pfam: PF00400; WD40; 8.
SMART; SM00320; WD40; 8.
DR PROSITE; PS00678; WD_REPEATS_1; 1.
DR PROSITE; PS00682; WD_REPEATS_2; 6.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
DR Repeat; WD repeat.
FT DOMAIN 14 18 HEPARIN-BINDING (POTENTIAL).
FT DOMAIN 71 77 POLY-GLO.
FT REPEAT 107 138 WD 1.
FT REPEAT 150 180 WD 2.
FT REPEAT 190 220 WD 3.
FT REPEAT 231 261 WD 4.
FT REPEAT 276 306 WD 5.
FT REPEAT 333 363 WD 6.
FT REPEAT 374 404 WD 7.
FT REPEAT 416 446 WD 8.
SQ SEQUENCE 452 AA; 49015 MW; DA1413D25EB236C0 CRC64;

Query Match 8.2%; Score 83; DB 1; Length 452;
Best Local Similarity 25.3%; Pred. No. 2.9;
Matches 42; Conservative 13; Mismatches 47; Indels 64; Gaps 9;

QY 68 WTVYHGAGTRTITASPKGPVTQMYTNVDKDLVGVQAPQGRSL-----TPCTGSSDLYLV 122
DB 197 WMEH-----PRAPVLLAGT-ADGNTWVKVPNGDCKTFOGPNCPATCGR----- 240
QY 123 TRHADVIPVRRR---GDSRGS-----LLSPRPISYLKSSG--GPLLCPA----- 162
DB 241 -----VLPDGRVAVGYEDGTIRINDLAKGSPFIRVLKGTGSHQGLTCVAAQDGLILT 295
QY 163 -----CHAVGIFR-----AAVSTRGVAKAVDFIPVESL 190
DB 296 GSVDCQAKLVSATTKGVGVFRPETAQPSQPSLGESESESNVESL 341

RESULT 13


```
RN SEQUENCE FROM N.A.
RX MEDLINE-92115700; PubMed-1731327;
RA Tsarev S.A., Emerson S.U., Reyes G.R., Tsareva T.S., Legters I.J.,
RA Malik I.A., Iqbal M., Purcell R.H.;
RT "Characterization of a prototype strain of hepatitis E virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:559-563(1992).
CC -1- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING
CC THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA
CC BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M80581; AAA5727.1; -.
DR InterPro; IPR004261; SP2.
DR Pfam; PF03014; SP2; 1.
KW Signal.
FT SIGNAL 1 22 BY SIMILARITY.
FT CHAIN 23 660 STRUCTURAL PROTEIN 2.
FT CHAIN 23 660 STRUCTURAL PROTEIN 2.
SQ SEQUENCE 660 AA; 70980 MW; 8085BC53CFB46FD3 CRC64;

Query Match 7.7%; Score 78.5; DB 1; Length 660;
Best Local Similarity 19.7%; Pred. No. 12;
Matches 45; Conservative 42; Mismatches 85; Indels 57; Gaps 10;

QY 10 VGRINLSGDTAYAAQTGEGTGKTSHTGDKNOV---EGEQIVYSTATQTFLA---T 61
Db 297 LGULDFALELEFNLTGNTNTRVSRYSSTARHRLRGADGTAELTTAATRFMKDLYFT 356
QY 62 SINGV-----LWTVYHGAGT-----RTIASPKG-PVTOMYTNV 93
Db 357 STNGVGEIGRGIALTLFNLADTLGLGPLEISSAGQLFYSRPVVSANGEPTVKLYTSV 416
QY 94 DKDLVGHQAQPGCSRLTPCTCGSSDLYLV---TRHADVIIPVRRGDSRG-SILSPRPISY 149
Db 417 ENA----QODKGIAPHDIDLGESRVVIQDYDNOHEQDRTPSPAPSPFVLRANDVLW 472
QY 150 LK-----GSSGGPLLCPAGHAVGIFRAAVSTRGVAKAVDFIPV 187
Db 473 LSLTAHYDOSTYCGSTGPNV--VDSVTIVNVATGNAQAVARSLDWTKV 519
```

Search completed: August 30, 2003, 19:13:50
Job time : 10.7567 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 19:00:22 ; Search time 37.5921 seconds
(without alignments)
1352.314 Million cell updates/sec

Title: US-09-965-594-22
Perfect score: 1016
Sequence: 1 MKKGSVVIVGRINLSGDTA.....VAKAVDFIPVESLETTMRSP 197

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTEMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriaph:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	862.5	84.9	4040	12 Q91FH8	Q91fh8 mucosal dis
2	848.5	83.5	3011	12 Q36579	Q36579 hepatitis c
3	844.5	83.1	2436	12 Q81756	Q81756 hepatitis c
4	844.5	83.1	3011	12 Q91FE5	Q91fe5 hepatitis c
5	844.5	83.1	3011	12 Q91ES8	Q91es8 hepatitis c
6	843.5	83.0	3011	12 Q03463	Q03463 hepatitis c
7	841.5	82.8	3011	12 Q36608	Q36608 hepatitis c
8	841.5	82.8	3015	12 Q9PWX5	Q9pwx5 hepatitis c
9	841.5	82.8	3015	12 Q9PWX5	Q9pwx5 hepatitis c
10	839	82.6	181	12 Q91RR8	Q91rr8 hepatitis c
11	839	82.6	181	12 Q91RR5	Q91rr5 hepatitis c
12	837	82.4	181	12 Q91RR5	Q91rr5 hepatitis c
13	837	82.4	181	12 Q91RR2	Q91rr2 hepatitis c
14	837	82.4	181	12 Q91RT9	Q91rt9 hepatitis c
15	836	82.3	181	12 Q91RR3	Q91rr3 hepatitis c
16	836	82.3	181	12 Q91RR4	Q91rr4 hepatitis c

17	836	82.3	181	12 Q91RS1	Q91rs1 hepatitis c
18	836	82.3	181	12 Q91RQ8	Q91rq8 hepatitis c
19	836	82.3	181	12 Q91RT1	Q91rt1 hepatitis c
20	836	82.3	181	12 Q91RR0	Q91rr0 hepatitis c
21	835.5	82.2	3011	12 Q36609	Q36609 hepatitis c
22	834	82.1	181	12 Q91RR6	Q91rr6 hepatitis c
23	834	82.1	181	12 Q91RS9	Q91rs9 hepatitis c
24	833	82.0	181	12 Q91RS3	Q91rs3 hepatitis c
25	832.5	81.9	3011	12 Q9DIT6	Q9dit6 hepatitis c
26	832	81.9	181	12 Q91RT4	Q91rt4 hepatitis c
27	832	81.9	181	12 Q91RS8	Q91rs8 hepatitis c
28	832	81.9	181	12 Q91RT3	Q91rt3 hepatitis c
29	832	81.9	181	12 Q91RS5	Q91rs5 hepatitis c
30	832	81.9	181	12 Q91RS7	Q91rs7 hepatitis c
31	832	81.9	181	12 Q91RT0	Q91rt0 hepatitis c
32	832	81.9	181	12 Q91RS2	Q91rs2 hepatitis c
33	831	81.8	181	12 Q91RS6	Q91rs6 hepatitis c
34	830.5	81.7	3010	12 Q9QP61	Q9qp61 hepatitis c
35	830	81.7	181	12 Q91RS4	Q91rs4 hepatitis c
36	829.5	81.6	3010	12 Q68533	Q68533 hepatitis c
37	829	81.6	181	12 Q91RR7	Q91rr7 hepatitis c
38	829	81.6	181	12 Q91RT6	Q91rt6 hepatitis c
39	829	81.6	3011	12 Q36610	Q36610 hepatitis c
40	828	81.5	181	12 Q91RT8	Q91rt8 hepatitis c
41	827.5	81.4	361	12 Q70818	Q70818 hepatitis c
42	827.5	81.4	361	12 Q70817	Q70817 hepatitis c
43	827	81.4	181	12 Q91RR9	Q91rr9 hepatitis c
44	826.5	81.3	3010	12 Q9DTE2	Q9dte2 hepatitis c
45	826.5	81.3	3010	12 Q99AU2	Q99au2 hepatitis c

ALIGNMENTS

RESULT 1

ID	Q91FH8	PRELIMINARY:	PRT: 4040 AA.
AC	Q91FH8;		
DR	01-OCT-2000 (TReMBLrel. 15, Created)		
DT	01-OCT-2000 (TReMBLrel. 15, Last sequence update)		
DT	01-MAR-2003 (TReMBLrel. 23, Last annotation update)		
DE	Genome polyprotein.		
OS	Mucosal disease virus.		
OC	Viruses; ssRNA positive-strand viruses, no DNA stage: Flaviviridae;		
OX	NCBI_TaxID=11099;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE-20323484; PubMed-10864644;		
RA	Lai V.C., Zhong W., Skelton A., Ingravallo P., Vassilev V.,		
RA	Donis R.O., Hong Z., Lau J.Y.;		
RT	*Generation and characterization of a hepatitis C virus NS3 protease-		
RT	dependent bovine viral diarrhea virus.*;		
RL	J. Virol. 74:6339-6347(2000).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RA	Lai V.C.H., Hong Z.;		
RL	Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF268278; AAF82566.1; -		
DR	HSSP; P26663; 1JXP.		
DR	MEOPRS; S31.001; -		
DR	InterPro; IPR000280; CDvir_endptsep80.		
DR	InterPro; IPR001410; DEAD.		
DR	InterPro; IPR004109; HCV_NS3.		
DR	InterPro; IPR002166; HCV_RdRP.		
DR	InterPro; IPR001650; Helicase_C.		
DR	InterPro; IPR001005; Myb_DNA_binding.		
DR	InterPro; IPR001568; RNase_T2.		
DR	InterPro; IPR007095; RNA_pol_DS_PS.		
DR	InterPro; IPR007094; RNA_pol_PSvir.		
DR	Pfam; PF02907; HCV_NS3; 1.		
DR	Pfam; PF00998; Viral_RdRP; 1.		


```

DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RdRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS0507; RDRP_POSITIVE; 1.
DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
FT NON_TER 1
FT 2436 2436
SQ SEQUENCE 2436 AA: 264734 MW: D7B9872900BE3125 CRC64;

Query Match 83.1%; Score 844.5; DB 12; Length 2436;
Best Local Similarity 82.4%; Pred. No. 5.1e-71;
Matches 168; Conservative 9; Mismatches 18; Indels 9; Gaps 1;

QY 3 KGSVVIVGRIN-----LSGDTAYAAQTRGEQGTOKTSHTGRDKNQVEGEVQIVST 53
Db 555 RRGREILLGPADGVSKGWRLLAPITAYAAQTRGLLGCITSLTGRDKNQVEGEVQIVST 614
QY 54 ATQTFLATLSINGLVTVYHGAGTRTIAASPKGPVTQMTNVDKLVGWAQPGSSRLTPCT 113
Db 615 AAQTFLATCINGVCVTVYHGAGTRTIAASPKGPVIQMTNVQDVLGWPAPQGSRLTPCT 674
QY 114 CGSSDLYLVTRHADVIPVRRGRDGRGSLSPRPISYLKSGSGGPLLCAGHAGVIFRAAV 173
Db 675 CGSSDLYLVTRHADVIPVRRGRDGRGSLSPRPISYLKSGSGGPLLCAGHAGVIFRAAV 734
QY 174 STRGVAKAVDFIPVESLETTMRSP 197
Db 735 CTRGVAKAVDFIPVENLETTMRSP 758

RESULT 4
ID Q9IFE5 PRELIMINARY; PRT; 3011 AA.
AC Q9IFE5;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21262212; PubMed=11369872;
RA Lanford R.E., Lee H., Chavez D., Guerra B., Brasky K.M.;
RT "Infectious cDNA clone of the hepatitis C virus genotype 1 prototype
sequence.";
RL J. Gen. Virol. 82:1291-1297(2001).
CC -/- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AF271632; AAF81759.1; -.
DR HSSP; P27958; 1AIV.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.

```

```

DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RdRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS00190; CYTOCHROME_C; 1.
DR PROSITE; PS0507; RDRP_POSITIVE; 1.
DR PROSITE; PS05021; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA: 327124 MW: 2489CE74AC864E58 CRC64;

Query Match 83.1%; Score 844.5; DB 12; Length 3011;
Best Local Similarity 82.4%; Pred. No. 6.7e-71;
Matches 168; Conservative 9; Mismatches 18; Indels 9; Gaps 1;

QY 3 KGSVVIVGRIN-----LSGDTAYAAQTRGEQGTOKTSHTGRDKNQVEGEVQIVST 53
Db 1005 RRGREILLGPADGVSKGWRLLAPITAYAAQTRGLLGCITSLTGRDKNQVEGEVQIVST 1064
QY 54 ATQTFLATLSINGLVTVYHGAGTRTIAASPKGPVTQMTNVDKLVGWAQPGSSRLTPCT 113
Db 1065 AAQTFLATCINGVCVTVYHGAGTRTIAASPKGPVIQMTNVQDVLGWPAPQGSRLTPCT 1124
QY 114 CGSSDLYLVTRHADVIPVRRGRDGRGSLSPRPISYLKSGSGGPLLCAGHAGVIFRAAV 173
Db 1125 CGSSDLYLVTRHADVIPVRRGRDGRGSLSPRPISYLKSGSGGPLLCAGHAGVIFRAAV 1184
QY 174 STRGVAKAVDFIPVESLETTMRSP 197
Db 1185 CTRGVAKAVDFIPVENLETTMRSP 1208

RESULT 5
ID Q9ELS8 PRELIMINARY; PRT; 3011 AA.
AC Q9ELS8;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=colonel;
RA Desai S.M., Devare S., Yamaguchi J.;
RT "Hepatitis C Virus.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -/- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AF290978; AAG02099.1; -.
DR HSSP; P27958; 1HEI.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR001410; DEAD.

```

```
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NSI.
DR InterPro: IPR002518; HCV_NS1.
DR InterPro: IPR004109; HCV_NS2.
DR InterPro: IPR000745; HCV_NS3.
DR InterPro: IPR001490; HCV_NS4a.
DR InterPro: IPR000745; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_NS5a.
DR InterPro: IPR001650; HCV_RdRP.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NSI; 1.
DR Pfam: PF01538; HCV_NS1; 1.
DR Pfam: PF02907; HCV_NS2; 1.
DR Pfam: PF01006; HCV_NS3; 1.
DR Pfam: PF01001; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RdRP; 1.
DR Pfam: PF0186062; HCV_NSI; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS00190; CYTOCHROME_C; 1.
DR PROSITE: PS05057; RDRP_POSITIVE; 1.
DR PROSITE: PS05052; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolyase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferrase; Transmembrane.
SQ SEQUENCE 3011 AA; 327107 MW; A6BECF5A3B3EE13F CRC64;

Query Match
Best Local Similarity 81.1%; Score 844.5; DB 12; Length 3011.
Matches 167; Conservative 11; Mismatches 17; Indels 9; Gaps 1;

QY 3 KKGSVIVGRIN-----LSGDTAYAQOTRGQGTQKTSHTGRDKNOVEGEVQIVST 53
DB 1005 RGQELILGPADGMVSKGWRLLAPITAYAQOTRGLGCIITSLTGRDKNOVEGEVQIVST 1064

QY 54 ATOTFLATSLNGVLVTVHAGTRTIASPKGPVQTMVNDKLVGWAQPGSRSLTPTCT 113
DB 1065 ATQTLATCINGVCTVTVHAGTRTIASPKGPVQTMVNDQDLVGPAPQCSRLTPTCT 1124

QY 114 CGSSDLYLVTRHADVIPVRRGDSRGLSLSPRPSYILKGGSGGGLLCPAGHVGIFRAAV 173
DB 1125 CGSSDLYLVTRHADVIPVRRGDSRGLSLSPRPSYILKGGSGGGLLCPAGHVGIFRAAV 1184

QY 174 STRGVAKAVDFIPVLESLETTMRSP 197
DB 1185 CTRGVAKAVDFIPVENLETTMRSP 1208

RESULT 6
Q03463 ID Q03463 PRELIMINARY; PRT: 3011 AA.
AC Q03463;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J1;
RX MEDLINE=91013116; PubMed=2170712;
RA Okamoto H., Okada S., Sugiyama Y., Yotsumoto S., Tanaka T.,
RA Yoshizawa H.;
```

```
RT "The 5'-terminal sequence of the hepatitis C virus genome.";
RL Jpn. J. Exp. Med. 60:167-177(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J1;
RX MEDLINE=92044440; PubMed=1658196;
RA Okamoto H., Okada S., Sugiyama Y., Kural K., Iizuka H., Machida A.,
RA Miyakawa Y., Mayumi M.;
RT "Nucleotide sequences of the genomic RNA of hepatitis C virus isolated
RT from a human carrier: comparison with reported isolates for conserved
RT and divergent regions.";
RL J. Gen. Virol. 72:2697-2704(1991).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J1;
RX MEDLINE=93117120; PubMed=1335573;
RA Okamoto H., Kanai N., Mishiro S.;
RT "Full-length nucleotide sequence of a Japanese hepatitis C virus
RT isolate (HC-J1) with high homology to USA isolates.";
RL Nucleic Acids Res. 20:6410-6410(1992).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J1;
RA Okamoto H.;
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J1;
RX MEDLINE=94174722; PubMed=7510436;
RA Mink M., Benichou S., Madaule P., Tiollais P., Prince A.,
RA Inchausti G.;
RT "Characterization and mapping of a B-cell immunogenic domain in
RT hepatitis C virus E2 glycoprotein using a yeast peptide library.";
RL Virology 200:246-255(1994).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: D10749; BAA01582.1; -.
DR HSSP: P27958; 1HE1.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NSI.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRP.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NSI; 1.
DR Pfam: PF01538; HCV_NS1; 1.
DR Pfam: PF02907; HCV_NS2; 1.
DR Pfam: PF01006; HCV_NS3; 1.
DR Pfam: PF01001; HCV_NS4a; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RdRP; 1.
DR ProDom: PD186062; HCV_NSI; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS05057; RDRP_POSITIVE; 1.
DR PROSITE: PS05052; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolyase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferrase; Transmembrane.
SQ SEQUENCE 3011 AA; 327112 MW; 97E9052C0250463B CRC64;
```

Query Match 83.0%; Score 843.5; DB 12; Length 3011;
Best Local Similarity 82.4%; Pred. No. 8.4e-71;
Matches 168; Conservative 8; Mismatches 19; Indels 9; Gaps 1;

QY 3 KGSVVIVGRIN-----LSGDTAYAQOTRGEQGTQKTSHTGRDKNQVEGEVQIVST 53
DB 1005 RRGQELLGPGADGMSKGRLLAPITAYAQOTRGLGCIITSLTGRDKNQVEGEVQIVST 1064

QY 54 ATOTFLATSLINGVLTVYHGAGTRTIASPKGPVTOMYTNVDKLVGMQAPGQSRSLTPCT 113
DB 1065 AAOTFLATSLINGVLTVYHGAGTRTIASPKGPVTOMYTNVDKLVGMQAPGQSRSLTPCT 1124

QY 114 CGSSDLYLVTRHADVIPVRRGRDGRGSLSPRPISYLGKSSGGPLLCPCAGHAGVIFRAAV 173
DB 1125 CGSSDLYLVTRHADVIPVRRGRDGRGSLSPRPISYLGKSSGGPLLCPCAGHAGVIFRAAV 1184

QY 174 STRGVAKAVDFIPVESLETTMRSP 197
DB 1185 CTRGVAKAVDFIPVESLETTMRSP 1208

RESULT 7
O36608 PRELIMINARY; PRT; 3011 AA.

AC O36608;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus strain H77.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=63746;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H77;
RX MEDLINE=97385173; PubMed=9238047;
RA Yanagi M., Purcell R.H., Emerson S.O., Bukh J.;
RT "Transcripts from a single full-length cDNA clone of hepatitis C virus
are infectious when directly transfected into the liver of a
chimpanzee.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743(1997).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL: AF011751; AAB67036.1; -;
DR HSSP: P27958; 1HEI.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRP.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.

DR Pfam: PF00998; Viral_RdRP; 1.
DR PRODom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS50507; RDRP_POSITIVE; 1.
DR PROSITE: PS50521; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolyase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA; 327112 MW; 0B75E6B81CB5C198 CRC64;

Query Match 82.8%; Score 841.5; DB 12; Length 3011;
Best Local Similarity 81.9%; Pred. No. 1.3e-70;
Matches 167; Conservative 10; Mismatches 18; Indels 9; Gaps 1;

QY 3 KGSVVIVGRIN-----LSGDTAYAQOTRGEQGTQKTSHTGRDKNQVEGEVQIVST 53
DB 1005 RRGQELLGPGADGMSKGRLLAPITAYAQOTRGLGCIITSLTGRDKNQVEGEVQIVST 1064

QY 54 ATOTFLATSLINGVLTVYHGAGTRTIASPKGPVTOMYTNVDKLVGMQAPGQSRSLTPCT 113
DB 1065 AAOTFLATSLINGVLTVYHGAGTRTIASPKGPVTOMYTNVDKLVGMQAPGQSRSLTPCT 1124

QY 114 CGSSDLYLVTRHADVIPVRRGRDGRGSLSPRPISYLGKSSGGPLLCPCAGHAGVIFRAAV 173
DB 1125 CGSSDLYLVTRHADVIPVRRGRDGRGSLSPRPISYLGKSSGGPLLCPCAGHAGVIFRAAV 1184

QY 174 STRGVAKAVDFIPVESLETTMRSP 197
DB 1185 CTRGVAKAVDFIPVESLETTMRSP 1208

RESULT 8
Q9PWX5 PRELIMINARY; PRT; 3015 AA.

AC Q9PWX5;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99420396; PubMed=10489358;
RA Yanagi M., Purcell R.H., Emerson S.O., Bukh J.;
RT "Hepatitis C virus: an infectious molecular clone of a second major
genotype (2a) and lack of viability of intertypic 1a and 2a
chimeras.";
RL Virology 262:250-263(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Bukh J.;
RL Submitted
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL: AF177040; AAF01182.1; -;
DR HSSP: AF177038; AAF01180.1; -;
DR HSSP: P27958; 1HEI.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRP.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.

```

DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR002129; Pyridoxal_deC.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; Helicase_C; 1.
DR Pfam: PF00998; Viral_RDRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXdc; 1.
DR PROSITE: PS00392; DDC_GAD_HDC_YDC; 1.
DR PROSITE: PS05057; RDRP_POSITIVE; 1.
DR PROSITE: PS05021; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3015 AA; 328159 MW; B7D23BC1F190663A CRC64;

Query Match 82.8%; Score 841.5; DB 12; Length 3015;
Best Local Similarity 81.9%; Pred. No. 1.3e-70;
Matches 167; Conservative 10; Mismatches 18; Indels 9; Gaps 1;

Qy 3 KGSVVIVGRIN-----LSGDTAYAOQTREGQGTQKTSHTGRDKNQVEGEVIVST 53
Db 1009 RRGQELLGPADGMVSKGWRLLAPITAYAOQTREGLLGCIITSLTGRDKNQVEGEVIVST 1068

Qy 54 ATQTFLATISINGVLWTVYHGAGTRTASPKGPVTQMTYNDKDLVGWQAPQGSRLTPCT 113
Db 1069 ATQTFLATISINGVLWTVYHGAGTRTASPKGPVTQMTYNDKDLVGWQAPQGSRLTPCT 1128

Qy 114 CGSSDLYLVTRHADVIPVRRGDSRGLSPRISYLGSGSGPLLCPCGAVGIFRAAV 173
Db 1129 CGSSDLYLVTRHADVIPVRRGDSRGLSPRISYLGSGSGPLLCPCGAVGIFRAAV 1188

Qy 174 STRGVAKAVDFIPVESLETTMRSP 197
Db 1189 CTRGVAKAVDFIPVENLGTMRSP 1212

RESULT 9
Q9PM09 PRELIMINARY; PRT; 3015 AA.
AC Q9PM09;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99420396; PubMed=10489358;
RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;
RT "Hepatitis C virus: an infectious molecular clone of a second major
RT genotype (2a) and lack of viability of intertypic 1a and 2a
RT chimeras."
RT Virology 262:250-263(1999).
RL [2]
RN [2]
RP SEQUENCE FROM N.A.
RA Bukh J.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

```

```

CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: AF177039; AAF01181.1; -.
DR EMBL: AF177037; AAF01179.1; -.
DR HSSP: P27958; 1HEI.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RDRP.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR002129; Pyridoxal_deC.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; Helicase_C; 1.
DR Pfam: PF00998; Viral_RDRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXdc; 1.
DR PROSITE: PS00392; DDC_GAD_HDC_YDC; 1.
DR PROSITE: PS05057; RDRP_POSITIVE; 1.
DR PROSITE: PS05021; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3015 AA; 328084 MW; E309F6318067D6CD CRC64;

Query Match 82.8%; Score 841.5; DB 12; Length 3015;
Best Local Similarity 81.9%; Pred. No. 1.3e-70;
Matches 167; Conservative 10; Mismatches 18; Indels 9; Gaps 1;

Qy 3 KGSVVIVGRIN-----LSGDTAYAOQTREGQGTQKTSHTGRDKNQVEGEVIVST 53
Db 1009 RRGQELLGPADGMVSKGWRLLAPITAYAOQTREGLLGCIITSLTGRDKNQVEGEVIVST 1068

Qy 54 ATQTFLATISINGVLWTVYHGAGTRTASPKGPVTQMTYNDKDLVGWQAPQGSRLTPCT 113
Db 1069 ATQTFLATISINGVLWTVYHGAGTRTASPKGPVTQMTYNDKDLVGWQAPQGSRLTPCT 1128

Qy 114 CGSSDLYLVTRHADVIPVRRGDSRGLSPRISYLGSGSGPLLCPCGAVGIFRAAV 173
Db 1129 CGSSDLYLVTRHADVIPVRRGDSRGLSPRISYLGSGSGPLLCPCGAVGIFRAAV 1188

Qy 174 STRGVAKAVDFIPVESLETTMRSP 197
Db 1189 CTRGVAKAVDFIPVENLGTMRSP 1212

RESULT 10
Q91RR8 PRELIMINARY; PRT; 181 AA.
AC Q91RR8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;

```

```

RN  SEQUENCE FROM N.A.
RC  STRAIN-Pt.1V;
RA  Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT  "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RL  Clinical Strains of the Hepatitis C Virus.";
DR  EMBL: AF369235; AAK54560.1; -
DR  InterPro: IPR004109; HCV_NS3.
DR  Pfam: PF02907; HCV_NS3; 1.
KW  Protease.
FT  NON_TER 1 1
FT  NON_TER 181 181
SQ  SEQUENCE 181 AA: 19130 MW: 85091869299B7C35 CRC64;

Query Match 82.6%; Score 839; DB 12; Length 181;
Best Local Similarity 92.7%; Pred. No. 5.5e-72;
Matches 165; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

QY 19 TAYAAQTRGEGQGTOKTSHTGRDNQVEGEVQIVSTATOTFLATSIINGVLWTVYHGAGTGT 78
DB 4 TAYAAQTRGLGCIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTYVHGAGTGT 63
QY 79 IASPKGPVTQMTYNDKDLVGWQAPQGSRLTPTCTGSSDLYLVTRHADVIPVRRGDSR 138
DB 64 IASPKGPVTQMTYNDKDLVGWQAPQGSRLTPTCTGSSDLYLVTRHADVIPVRRGDSR 123
QY 139 GSLLSPRPISYLGSSGGPLCPAGHAVGIFRAAVSTRGVAKAVDFIPVESLETTMRS 196
DB 124 GSLLSPRPISYLGSSGGPLCPAGHAVGIFRAAVSTRGVAKAVDFIPVESLETTMRS 181

RESULT 11
Q91RT5 PRELIMINARY; PRT; 181 AA.
ID Q91RT5
AC Q91RT5;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Pt.1V;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RL Clinical Strains of the Hepatitis C Virus.";
DR EMBL: AF369218; AAK54543.1; -
DR InterPro: IPR004109; HCV_NS3.
DR Pfam: PF02907; HCV_NS3; 1.
KW Protease.
FT NON_TER 1 1
FT NON_TER 181 181
SQ SEQUENCE 181 AA: 19130 MW: 85D91869299B7C35 CRC64;

Query Match 82.6%; Score 839; DB 12; Length 181;
Best Local Similarity 92.7%; Pred. No. 5.5e-72;
Matches 165; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

QY 19 TAYAAQTRGEGQGTOKTSHTGRDNQVEGEVQIVSTATOTFLATSIINGVLWTVYHGAGTGT 78
DB 4 TAYAAQTRGLGCIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTYVHGAGTGT 63
QY 79 IASPKGPVTQMTYNDKDLVGWQAPQGSRLTPTCTGSSDLYLVTRHADVIPVRRGDSR 138
DB 64 IASPKGPVTQMTYNDKDLVGWQAPQGSRLTPTCTGSSDLYLVTRHADVIPVRRGDSR 123
QY 139 GSLLSPRPISYLGSSGGPLCPAGHAVGIFRAAVSTRGVAKAVDFIPVESLETTMRS 196
DB 124 GSLLSPRPISYLGSSGGPLCPAGHAVGIFRAAVSTRGVAKAVDFIPVESLETTMRS 181

RESULT 12
Q91RR5 PRELIMINARY; PRT; 181 AA.
ID Q91RR5
AC Q91RR5;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Pt.3U;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RL Clinical Strains of the Hepatitis C Virus.";
DR EMBL: AF369238; AAK54563.1; -
DR InterPro: IPR004109; HCV_NS3.
DR Pfam: PF02907; HCV_NS3; 1.
KW Protease.
FT NON_TER 1 1
FT NON_TER 181 181
SQ SEQUENCE 181 AA: 19084 MW: 3B5E8161F2100A72 CRC64;

Query Match 82.4%; Score 837; DB 12; Length 181;
Best Local Similarity 92.1%; Pred. No. 8.5e-72;
Matches 164; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 19 TAYAAQTRGEGQGTOKTSHTGRDNQVEGEVQIVSTATOTFLATSIINGVLWTVYHGAGTGT 78
DB 4 TAYAAQTRGLGCIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTYVHGAGTGT 63
QY 79 IASPKGPVTQMTYNDKDLVGWQAPQGSRLTPTCTGSSDLYLVTRHADVIPVRRGDSR 138
DB 64 IASPKGPVTQMTYNDKDLVGWQAPQGSRLTPTCTGSSDLYLVTRHADVIPVRRGDSR 123
QY 139 GSLLSPRPISYLGSSGGPLCPAGHAVGIFRAAVSTRGVAKAVDFIPVESLETTMRS 196
DB 124 GSLLSPRPISYLGSSGGPLCPAGHAVGIFRAAVSTRGVAKAVDFIPVESLETTMRS 181

RESULT 13
Q91RR2 PRELIMINARY; PRT; 181 AA.
ID Q91RR2
AC Q91RR2;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Pt.4V;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RL Clinical Strains of the Hepatitis C Virus.";
DR EMBL: AF369241; AAK54566.1; -
DR InterPro: IPR004109; HCV_NS3.
DR Pfam: PF02907; HCV_NS3; 1.
KW Protease.
FT NON_TER 1 1
FT NON_TER 181 181
SQ SEQUENCE 181 AA: 19123 MW: 1CAE817345ED809D CRC64;
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: August 30, 2003, 19:18:33 ; Search time 2560.57 Seconds
(without alignments)
3147.423 Million cell updates/sec

Title: US-09-965-594-22

Perfect score: 1016

Sequence: 1 MKKGGVVIVGRINLSGDTA.....VAKAVDFIPVESLETTMRSP 197

Scoring table: BLOSUM62 Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xl
-Q=/cqn2_1/USPTO_spool/US0965594/runat_29082003_151919_28310/app_query.fasta_1.2872
-DB=GenEmbl -OFMT=faastp -SUFFIX=rge -MINMATCH=0.1 -LOOPL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NOR=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0965594.ecgn_1_14686/runat_29082003_151919_28310 -NCPU=6 -ICPU=3
-NO_MMAPP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*
1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vi:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.mu:*
20: em.om:*
21: em.or:*
22: em.ov:*
23: em.pat:*
24: em.ph:*
25: em.pl:*
26: em.ro:*
27: em.sts:*
28: em.un:*

29: em.vi:*
30: em.htg_hum:*
31: em.htg_inv:*
32: em.htg_other:*
33: em.htg_mus:*
34: em.htg_pln:*
35: em.htg_rnd:*
36: em.htg_man:*
37: em.htg_vrt:*
38: em.sy:*
39: em.htgo_hum:*
40: em.htgo_mus:*
41: em.htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Query Length	DB ID	Description
1	882.5	86.9	12734	6	AR179057 Sequence
2	875.5	86.2	1998	6	AR145264 Sequence
3	872.5	85.9	1998	6	AR145268 Sequence
4	871.5	85.8	1998	6	AR145263 Sequence
5	868.5	85.5	651	6	AR145254 Sequence
6	868.5	85.5	1998	6	AR145267 Sequence
7	867.5	85.4	1998	6	AR145262 Sequence
8	865.5	85.2	651	6	AR145258 Sequence
9	864.5	85.1	651	6	AR145253 Sequence
10	864.5	85.1	1998	6	AR145266 Sequence
11	863.5	85.0	1998	6	AR145261 Sequence
12	862.5	85.0	2016	6	AR145269 Sequence
13	862.5	84.9	12734	14	AF268278 Pestivirus
14	861.5	84.8	651	6	AR145257 Sequence
15	860.5	84.7	651	6	AR145252 Sequence
16	860.5	84.7	1998	6	AR145265 Sequence
17	860.5	84.7	2016	6	AR145270 Sequence
18	860	84.6	648	6	AR145274 Sequence
19	858	84.4	648	6	AR145272 Sequence
20	857.5	84.4	651	6	AR145256 Sequence
21	857.5	84.4	651	6	AR145260 Sequence
22	856.5	84.3	651	6	AR145251 Sequence
23	856	84.3	648	6	AR145273 Sequence
24	854	84.1	648	6	AR145271 Sequence
25	853.5	84.0	651	6	AR145255 Sequence
26	853.5	84.0	651	6	AR145259 Sequence
27	851	83.8	8157	6	AR127810 Sequence
28	851	83.8	8157	6	BD081911 Hepatitis
29	849	83.6	1932	6	AR127809 Sequence
30	849	83.6	1932	6	BD081910 Hepatitis
31	848.5	83.5	9646	6	AR110828 Sequence
32	848.5	83.5	9646	6	BD069982 Functiona
33	848.5	83.5	9646	14	AF009606 Hepatitis
34	848.5	83.5	12980	6	AR110831 Sequence
35	848.5	83.5	12980	6	BD069985 Functiona
36	844.5	83.1	5360	6	AR118686 Sequence
37	844.5	83.1	5360	6	I06434 Sequence 48
38	844.5	83.1	5360	6	I09328 Sequence 8
39	844.5	83.1	6785	6	AR118692 Sequence
40	844.5	83.1	6785	6	I06440 Sequence 54
41	844.5	83.1	6785	6	I09329 Sequence 10
42	844.5	83.1	7310	6	AR118696 Sequence
43	844.5	83.1	7310	6	I09331 Sequence 15
44	844.5	83.1	7310	14	HPCPOLYP
45	844.5	83.1	8316	6	AR118703 Sequence

ALIGNMENTS

ARI179057
LOCUS ARI179057 12734 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6326137.
ACCESSION ARI179057
VERSION ARI179057.1 GI:20220612
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 12734)
AUTHORS Hong, Z., Lai, V. C. H. and Lau, J. Y. N.
TITLE Hepatitis C virus protease-dependent chimeric pestivirus
JOURNAL Patent: US 6326137-A 1 04-DEC-2001;
FEATURES Location/Qualifiers
source 1..12734
BASE COUNT 4032 a 2604 c 3295 g 2803 t
ORIGIN
Alignment Scores:
Pred. No.: 3 77e-64 Length: 12734
Score: 882.50 Matches: 176
Percent Similarity: 92.31% Conservative: 4
Best Local Similarity: 90.26% Mismatches: 12
Query Match: 86.86% Indels: 3
DB: 6 Gaps: 1
US-09-965-594-22 (1-197) x ARI179057 (1-12734)
Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
Db 413 GGTAGTGTGTATTTGTTAGTAAATTTTATCTGTTAGTGTAGTATCATCAGCGGTAC 472
Qy 22 AlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAspLys 41
Db 473 GCGGACAGACGAGAGGCGCTCTAGGGTGAAGATCACCATCTGACTGCGCGGACAAA 532
Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
Db 533 AACCAAGTGGAGGGTGAGTCCAGATCGTCACTGCTACCCAAACCTTCTGGCAACG 592
Qy 62 SerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
Db 593 TGCATCAATGGGGTATGTGGACTGCTACCCAGGGCGGCGGACGAGACCATCGCATCA 652
Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
Db 653 CCCAAGGTCTGTATCCAGATGATACCAATGTGGACCAAGACCTTGTGGCTGGCC 712
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
Db 713 GCTCTCAAGTTCCTCGCTCATGTACACCTCGACCTCGGCTCTCTCGGACCTTTACCTG 772
Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
Db 773 GTTACGAGCACCGCAGCTCATTCCTGCGCGCGGCGAGGTGATAGAGGGTAGCCTG 832
Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
Db 833 CTTTCGCCCGGCCATTTCTACCTAAAGGCTCTCTCGGGGGTCCGCTGTGTGCCCC 892
Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
Db 893 GCGGGACACCGGTGGGCTATTTCAGGCGCGCGGTGTCACCGGTGGAGTGCCCAAGCGC 952
Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196
Db 953 GTGGACTTTATCCCTGTGGAGAACCTAGACAAACCATGAGATCC 997
RESULT 2
ARI15264 1998 bp DNA linear PAT 08-AUG-2001
LOCUS
DEFINITION Sequence 105 from patent US 6211338.

ACCESSION ARI45264
VERSION ARI45264.1 GI:15107131
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1998)
AUTHORS Malcolm, B. A., Taremi, S. Shane., Weber, P. C. and Yao, N.
TITLE Single-chain recombinant complexes of hepatitis C virus NS3
protease and NS4A cofactor peptide
JOURNAL Patent: US 6211338-A 105 03-APR-2001;
FEATURES Location/Qualifiers
source 1..1998
BASE COUNT 411 a 595 c 569 g 423 t
ORIGIN
Alignment Scores:
Pred. No.: 1 73e-64 Length: 1998
Score: 875.50 Matches: 167
Percent Similarity: 92.86% Conservative: 15
Best Local Similarity: 85.20% Mismatches: 11
Query Match: 86.17% Indels: 3
DB: 6 Gaps: 1
US-09-965-594-22 (1-197) x ARI45264 (1-1998)
Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
Db 64 GGTCTGTGTATTGTTGGTAGAATTTATCTGTTAGTGTAGTATCAGCGCTAC 123
Qy 22 AlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAspLys 41
Db 124 TCCCAACAGACGCGGGCGCTACTTGGTTGCAAGAAGACTAGCTTTACAGGCGGACAA 183
Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
Db 184 AACCAAGTGGAGGGTGGAGGTTTCCAGTGGTTTCCACCGCAACAATCTCTCTGCGGACC 243
Qy 62 SerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
Db 244 TCGCTCAACGCGGTGTGGACCGTTTACCATGTGTGGCTCAAGACCTTACCGGC 303
Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
Db 304 CCAAGGGCCCAATACCCAGATGATACATAATGTGGACAGGACCTCTCGGCTGGCAG 363
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
Db 364 GCGCCCCCGGGGCGGTCTCTTGACACCATGCACCTGTGGCAGCTCAGACCTTTACTTG 423
Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
Db 424 GTCACGAGACATGCTCAGCTCATTCGGTGGCGCGCGGCGGACAGTAGGGGAGCGCTG 483
Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
Db 484 CTCTCCCCAGGCGCTCTCTCTACTTGAAGGGCTCTCTGGGTGGTCCACTGCTCTGCCCT 543
Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
Db 544 TCGGGGACGCTGTGGCATCTTCCGGGTGCGGTATGCACCGGGGGGTTCGGAAGCGG 603
Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
Db 604 GTGGACTTTGTGCGCGTAGAGTCCATGGAACACTACTATGCGGCTCTCCG 651
RESULT 3
ARI45268 1998 bp DNA linear PAT 08-AUG-2001
LOCUS
DEFINITION Sequence 109 from patent US 6211338.
ACCESSION ARI45268
VERSION ARI45268.1 GI:15107135

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1998)

AUTHORS Malcolm,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.

TITLE Single-chain recombinant complexes of hepatitis C virus NS3

JOURNAL protease and NS4A cofactor peptide

PATENT: US 6211338-A 109 03-APR-2001;

FEATURES Location/Qualifiers

source

1..1998

/organism="unknown"

BASE COUNT 411 a 595 c 569 g 423 t

ORIGIN

Alignment Scores:

Pred. No.: 3 09e-64 Length: 1998

Score: 872.50 Matches: 166

Percent Similarity: 92.86% Conservatives: 16

Best Local Similarity: 84.69% Mismatches: 11

Query Match: 85.88% Indels: 3

DB: 6 Gaps: 1

US-09-965-594-22 (1-197) x ARI45268 (1-1998)

QY 5 GlySerValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21

Dy 64 GGTCTGTTGTTATTTGGTAGAATTTATCTGTTAGTATCATCGGCTAC 123

QY 22 AlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAspLys 41

Dy 124 TCCCAACAGACGGGGGCTACTTGGTTCAGAAAGACTAGCTTACAGCGGGGACAAG 183

QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61

Dy 184 AACCAAGTCGAGGAGAGGTTTCAGTGTTCACCGCAACACATCTCTCTGGCGACC 243

QY 62 SerIleAsnGlyValLeuThrThrValThrHisGlyAlaGlyThrArgThrIleAlaSer 81

Dy 244 TCGCTCAACGGGCTGTGTTGGACCTTTACCATGTGCTGCTCAAGACCTTAGCCCGC 303

QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101

Dy 304 CCNAGGGGCCAATCACCAGATGTACACTATGTGGACGAGCTCTGGCTGGCAG 363

QY 102 AlaProGlnCysArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121

Dy 364 GCGCCCGCGGGCGGCTCTCTTGACCATGACCTGTGGCAGCTCAGACCTTTACTTG 423

QY 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141

Dy 424 GTCACGAGACATGCTGACGTCTATCCGGTGGCGGGGGGCGACAGTAGGGGAGCCTG 483

QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161

Dy 484 CTCCTCCCGAGGCTCTCTCTACTTGAAGGCTCTGCTGGTGGTCCACTGCTCTGCCCT 543

QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181

Dy 544 TCGGGGACGCTGTGGGCTCTCTCCGGCTGCGGTATGCACCGGGGGGTTTCGAAGGCG 603

QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197

Dy 604 GTGGACTTTGTGCCCGTAGAGTCCATGGAACACTACTATCGGCTCTCCG 651

RESULT 4

ARI45263

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

Sequence 104 from patent US 6211338.

ARI45263

ARI45263.1 GI:15107130

Unknown.

ORGANISM

Unknown.

Unclassified.

REFERENCE 1 (bases 1 to 1998)

AUTHORS Malcolm,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.

TITLE Single-chain recombinant complexes of hepatitis C virus NS3

JOURNAL protease and NS4A cofactor peptide

PATENT: US 6211338-A 104 03-APR-2001;

FEATURES Location/Qualifiers

source

1..1998

/organism="unknown"

BASE COUNT 410 a 596 c 568 g 424 t

ORIGIN

Alignment Scores:

Pred. No.: 3 75e-64 Length: 1998

Score: 871.50 Matches: 167

Percent Similarity: 92.35% Conservatives: 14

Best Local Similarity: 85.20% Mismatches: 12

Query Match: 85.78% Indels: 3

DB: 6 Gaps: 1

US-09-965-594-22 (1-197) x ARI45263 (1-1998)

QY 5 GlySerValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21

Dy 64 GGTCTGTTGTTATTTGGTAGAATTTATCTGTTAGTATCATCGGCTAC 123

QY 22 AlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAspLys 41

Dy 124 TCCCAACAGACGGGGGCTACTTGGTTCAGAAAGACTAGCTTACAGCGGGGACAAG 183

QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61

Dy 184 AACCAAGTCGAGGAGAGGTTTCAGTGTTCACCGCAACACATCTCTCTGGCGACC 243

QY 62 SerIleAsnGlyValLeuThrThrValThrHisGlyAlaGlyThrArgThrIleAlaSer 81

Dy 244 TCGCTCAACGGGCTGTGTTGGACCTTTACCATGTGCTGCTCAAGACCTTAGCCCGC 303

QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101

Dy 304 CCNAGGGGCCAATCACCAGATGTACACTATGTGGACGAGCTCTGGCTGGCAG 363

QY 102 AlaProGlnCysArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121

Dy 364 GCGCCCGCGGGCGGCTCTCTTGACCATGACCTGTGGCAGCTCAGACCTTTACTTG 423

QY 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141

Dy 424 GTCACGAGACATGCTGACGTCTATCCGGTGGCGGGGGGCGACAGTAGGGGAGCCTG 483

QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161

Dy 484 CTCCTCCCGAGGCTCTCTCTACTTGAAGGCTCTGCTGGTGGTCCACTGCTCTGCCCT 543

QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181

Dy 544 TCGGGGACGCTGTGGGCTCTCTCCGGCTGCGGTATGCACCGGGGGGTTTCGAAGGCG 603

QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197

Dy 604 GTGGACTTTGTGCCCGTAGAGTCCATGGAACACTACTATCGGCTCTCCG 651

RESULT 5

ARI45254

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

Sequence 95 from patent US 6211338.

ARI45254

ARI45254.1 GI:15107121

Unknown.

Unknown.

Unknown.

Unknown.

Unknown.

Unknown.

Unknown.

Unknown.

Unknown.

Unknown.

Unknown.

REFERENCE 1 (bases 1 to 651)
 Malcolm.B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.
 TITLE Single-chain recombinant complexes of hepatitis C virus NS3
 protease and NS4A cofactor peptide
 JOURNAL Patent: US 6211338-A 95 03-APR-2001;
 FEATURES Location/Qualifiers

BASE COUNT 120 a 187 c 200 g 144 t
 ORIGIN 1..651
 /organism="unknown"

Alignment Scores:
 Pred. No.: 1.84e-64 Length: 651
 Score: 868.50 Matches: 166
 Percent Similarity: 92.82% Conservative: 15
 Best Local Similarity: 85.13% Mismatches: 11
 Query Match: 85.48% Indels: 3
 DB: 6 Gaps: 1

US-09-965-594-22 (1-197) x ARI45254 (1-651)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
 |||||||
 Db 64 GTTCTGTTGTTATGTTGGTAGAATATTATCTGCTAGTGTAGTATCAGCGCTAC 123
 QY 22 AlaGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAspLys 41
 |||||||
 Db 124 TCCCAACAGACGCGGGCCCTACTTGGTTGCAAGAGACTAGCCTTACAGCCGGGACAAG 183
 QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
 |||||||
 Db 184 ACCAGTCCAGGAGAGGTTTCAGTGGTTCCACCGCAACACATCCTTCTGGCGACC 243
 QY 62 SerIleAsnGlyValLeuThrValThrValHisGlyAlaGlyThrArgThrIleAlaSer 81
 |||||||
 Db 244 TCGTCAACGGCGTGTCTGGACCGTTTACATGGTGTGGCTCAAAGACCTTAGCGCGC 303
 QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
 |||||||
 Db 304 CCAAGGGGCAATCACCCAGATGTACATAATGTGGACAGGACCTCGCTGGCGCAG 363
 QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
 |||||||
 Db 364 CGCCCCCGGGCGCGTCTCTTGACACCATGTCACCTCTGGCAGCTCAGACCTTTACTTG 423
 QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
 |||||||
 Db 424 GTACAGACATGTCACGTCATTCCGGTGGCGGGCGGCGGAGGAGGAGGAGGAGGAGG 483
 QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
 |||||||
 Db 484 CTCTCCCCAGGCGCTCTCTCTACTTTGAAGGCTCTGCTGGTGTCTGCTGCTGCTGCT 543
 QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
 |||||||
 Db 544 TCGGGGACGCTGTGGGCATCTTCGGGCTGCGGTATGCACCGGGGGGTTCGGAAGCGC 603
 QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196
 |||||||
 Db 604 GTGGACTTTGGCCGTAGAGTCCATGGAATACTATATGCGGTCT 648

RESULT 6
 ARI45267
 LOCUS ARI45267 1998 bp DNA linear PAT 08-AUG-2001
 DEFINITION Sequence 108 from patent US 6211338.
 ACCESSION ARI45267
 VERSION ARI45267.1 GI:15107134
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 Unclassified.
 REFERENCE 1 (bases 1 to 1998)
 AUTHORS Malcolm.B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.

TITLE Single-chain recombinant complexes of hepatitis C virus NS3
 protease and NS4A cofactor peptide
 JOURNAL Patent: US 6211338-A 108 03-APR-2001;
 FEATURES Location/Qualifiers

BASE COUNT 410 a 596 c 568 g 424 t
 ORIGIN 1..1998
 /organism="unknown"

Alignment Scores:
 Pred. No.: 6.72e-64 Length: 1998
 Score: 868.50 Matches: 166
 Percent Similarity: 92.35% Conservative: 15
 Best Local Similarity: 84.69% Mismatches: 12
 Query Match: 85.48% Indels: 3
 DB: 6 Gaps: 1

US-09-965-594-22 (1-197) x ARI45267 (1-1998)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
 |||||||
 Db 64 GTTCTGTTGTTATGTTGGTAGAATATTATCTGCTAGTGTAGTATCAGCGCTAC 123
 QY 22 AlaGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAspLys 41
 |||||||
 Db 124 TCCCAACAGACGCGGGCCCTACTTGGTTGCAAGAGACTAGCCTTACAGCGCGGACAAG 183
 QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
 |||||||
 Db 184 ACCAGTCCAGGAGAGGTTTCAGTGGTTCCACCGCAACACATCCTTCTGGCGACC 243
 QY 62 SerIleAsnGlyValLeuThrValThrValHisGlyAlaGlyThrArgThrIleAlaSer 81
 |||||||
 Db 244 TCGTCAACGGCGTGTCTGGACCGTTTACATGGTGTGGCTCAAAGACCTTAGCGCGC 303
 QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
 |||||||
 Db 304 CCAAGGGGCAATCACCCAGATGTACATAATGTGGACAGGACCTCGCTGGCGGAG 363
 QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
 |||||||
 Db 364 CGCCCCCGGGCGCGTCTCTTGACACCATGTCACCTCTGGCAGCTCAGACCTTTACTTG 423
 QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
 |||||||
 Db 424 GTACAGACATGTCACGTCATTCCGGTGGCGGGCGGCGGAGGAGGAGGAGGAGGAGG 483
 QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
 |||||||
 Db 484 CTCTCCCCAGGCGCTCTCTCTACTTTGAAGGCTCTGCTGGTGTCTGCTGCTGCTGCT 543
 QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
 |||||||
 Db 544 TCGGGGACGCTGTGGGCATCTTCGGGCTGCGGTATGCACCGGGGGGTTCGGAAGCGC 603
 QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
 |||||||
 Db 604 GTGGACTTTGGCCGTAGAGTCCATGGAATACTATATGCGGTCTCCG 651

RESULT 7
 ARI45262
 LOCUS ARI45262 1998 bp DNA linear PAT 08-AUG-2001
 DEFINITION Sequence 103 from patent US 6211338.
 ACCESSION ARI45262
 VERSION ARI45262.1 GI:15107129
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 Unclassified.
 REFERENCE 1 (bases 1 to 1998)
 AUTHORS Malcolm.B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.
 TITLE Single-chain recombinant complexes of hepatitis C virus NS3
 protease and NS4A cofactor peptide

JOURNAL Patent: US 6211338-A 103 03-APR-2001;
 FEATURES Location/Qualifiers
 source 1..1998
 /organism="unknown"

BASE COUNT 410 a 596 c 568 g 424 t
 ORIGIN

Alignment Scores:
 Pred. No.: 8,166-64 Length: 1998
 Score: 867.50 Matches: 166
 Percent Similarity: 92.35% Conservative: 15
 Best Local Similarity: 84.68% Mismatches: 12
 Query Match: 85.38% Indels: 3
 DB: 6 Gaps: 1

US-09-965-594-22 (1-197) x ARI45262 (1-1998)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
 DB 64 GGTTCGTGTTATTTGTTGTAAGATTATTTATCTGTAGTGTAGTATCAGCGCTAC 123
 QY 22 AlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAspLys 41
 DB 124 TCCCAACACAGCGGGGCTACTTGGTTGCAAGATCACTAGCTTACAGCGCGGACAG 183
 QY 42 AsnGlnValGluGlyValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
 DB 184 AACCAAGTCGAGGAGAGGTTCCAGGTGGTTTCCACCGCAACAAATCCTTCTCGCGACC 243
 QY 62 SerIleAsnGlyValLeuTyrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
 DB 244 TCGCTCAACGGCGTGTGGACCGTTTACCATGGTGTGGTCAAGACCTTAGCCGGC 303
 QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101
 DB 304 CCAAGGGGCCAATCACCAGATGTACACTATGTGGACAGACCTCGTCGGCTGGCAG 363
 QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
 DB 364 GCGCCCCCGGGCGCGTTCCTTGACACCATGCACCTGTGGCAGCTCAGACCTTTACTTG 423
 QY 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141
 DB 424 GTCAGAGACATGCTGACGTCATCCGGTGGCGCGGGGCGACAGTAGGGGAGCCTG 483
 QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyProLeuLeuCysPro 161
 DB 484 CTCCTCCCGCAGGCTCTCTCTACTTGAAGGCTCTGCTGGTGGTCCACTGCTCTGCCCT 543
 QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
 DB 544 TCGGGCACGCTGTGGGCATCTCCGGGCTCCGATGCACCGGGGGGTTGCGAAGCG 603
 QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
 DB 604 GTGGACTTTGTGCCGTAGAGTCCATGGAACTACTATCGGCTCTCCG 651

RESULT 8
 ARI45258
 LOCUS 651 bp DNA linear PAT 08-AUG-2001
 DEFINITION Sequence 99 from patent US 6211338.
 ACCESSION ARI45258
 VERSION ARI45258.1 GI:15107125
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 Unclassified.

REFERENCE 1 (bases 1 to 651)
 AUTHORS Malcolim,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.
 TITLE Single-chain recombinant complexes of hepatitis C virus NS3
 protease and NS4A cofactor peptide
 JOURNAL Patent: US 6211338-A 99 03-APR-2001;
 FEATURES Location/Qualifiers

source 1..651
 BASE COUNT 120 a 187 c 200 g 144 t
 ORIGIN

Alignment Scores:
 Pred. No.: 3,296-64 Length: 651
 Score: 865.50 Matches: 165
 Percent Similarity: 92.82% Conservative: 16
 Best Local Similarity: 84.62% Mismatches: 11
 Query Match: 85.19% Indels: 3
 DB: 6 Gaps: 1

US-09-965-594-22 (1-197) x ARI45258 (1-651)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
 DB 64 GGTTCGTGTTATTTGTTGTAAGATTATTTATCTGTAGTGTAGTATCAGCGCTAC 123
 QY 22 AlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAspLys 41
 DB 124 TCCCAACACAGCGGGGCTACTTGGTTGCAAGAAAGACTAGCCTTACAGCGCGGACAG 183
 QY 42 AsnGlnValGluGlyValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
 DB 184 AACCAAGTCGAGGAGAGGTTCCAGGTGGTTTCCACCGCAACAAATCCTTCTCGCGACC 243
 QY 62 SerIleAsnGlyValLeuTyrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
 DB 244 TCGCTCAACGGCGTGTGGACCGTTTACCATGGTGTGGTCAAGACCTTAGCCGGC 303
 QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101
 DB 304 CCAAGGGGCCAATCACCAGATGTACACTATGTGGACAGACCTCGTCGGCTGGCAG 363
 QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
 DB 364 GCGCCCCCGGGCGCGTTCCTTGACACCATGCACCTGTGGCAGCTCAGACCTTTACTTG 423
 QY 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141
 DB 424 GTCAGAGACATGCTGACGTCATCCGGTGGCGCGGGGCGACAGTAGGGGAGCCTG 483
 QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyProLeuLeuCysPro 161
 DB 484 CTCCTCCCGCAGGCTCTCTCTACTTGAAGGCTCTGCTGGTGGTCCACTGCTCTGCCCT 543
 QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
 DB 544 TCGGGCACGCTGTGGGCATCTCCGGGCTCCGATGCACCGGGGGGTTGCGAAGCG 603
 QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196
 DB 604 GTGGACTTTGTGCCGTAGAGTCCATGGAACTACTATCGGCTCT 648

RESULT 9
 ARI45253
 LOCUS 651 bp DNA linear PAT 08-AUG-2001
 DEFINITION Sequence 94 from patent US 6211338.
 ACCESSION ARI45253
 VERSION ARI45253.1 GI:15107120
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 Unclassified.

REFERENCE 1 (bases 1 to 651)
 AUTHORS Malcolim,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.
 TITLE Single-chain recombinant complexes of hepatitis C virus NS3
 protease and NS4A cofactor peptide
 JOURNAL Patent: US 6211338-A 94 03-APR-2001;
 FEATURES Location/Qualifiers
 source 1..651
 /organism="unknown"

BASE COUNT 119 a 188 c 199 g 145 t
ORIGIN

Alignment Scores:

Pred. No.: 4e-64 Length: 651
Score: 864.50 Matches: 166
Percent Similarity: 92.31% Conservative: 12
Best Local Similarity: 85.13% Mismatches: 3
Query Match: 85.09% Indels: 1
DB: 6 Gaps: 1

US-09-965-594-22 (1-197) x AR145253 (1-651)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
|||||
Db 64 GGTTCGTGTTATTGTTGTAGTAATATTTATCTGTTAGTGTAGTATCATCGGCGCTAC 123
|||||
QY 22 AlaGlnGlnThrArgGlyGluGlnGlnThrGlnLysThrSerHisThrGlyArgAspLys 41
:::|||||
Db 124 TCCCAACAGACGCGGGGCGCTACTTGGTTCATCAAGACTAGCCTTACAGGCGGGGACAG 183
|||||
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
|||||
Db 184 AACAGGTCGAGGAGAGGTTTCAGTGGTTTCCACCGCAACACAACTCCTTCTGGCGACC 243
|||||
QY 62 SerIleAsnGlyValLeuThrPheValThrHisGlyAlaGlyThrArgThrIleAlaSer 81
:::|||||
Db 244 TGGGTCAACGCGGTGTTGGACCGTTTACCATGGTGTGGCTCAAAAGACCTTAGCGCGC 303
|||||
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101
|||||
Db 304 CCAAGGGGCGCAATCCACAGATGTACATAATGTGGACAGACCTCGTGGCGTGGCAG 363
|||||
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
|||||
Db 364 CGCGCCCGCGGCGCGTTCCTTGACACCATGTCACCTGTGGCAGCTCAGACCTTTACTTG 423
|||||
QY 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141
|||||
Db 424 GTACAGAGACATCTGACGTCTCCGTGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCG 483
|||||
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
|||||
Db 484 CTCTCCCGCAGGCGTGTCTCTACTTGAAGGGCTCTCGGGTGTCTCCACTGCTCTGCCCT 543
|||||
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
:::|||||
Db 544 TCGGGGCACGCTGTGGCATCTTCGGGCTGCGGTATGCACCGGGGGGTTCGGAAGGCG 603
|||||
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196
|||||
Db 604 GTGGACTTGTGCCCGTAGAGTCCATGGAAACTACTATGCGGTCT 648

RESULT 10
AR145266
LOCUS AR145266 1998 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 107 from patent US 6211338.
ACCESSION AR145266
VERSION AR145266.1 GI:15107133
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.

REFERENCE
Unclassified.
1 (bases 1 to 1998)
AUTHORS Malcolm,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.
TITLE Single-chain recombinant complexes of hepatitis C virus NS3
protease and NS4A cofactor peptide
JOURNAL Patent: US 6211338-A 107 03-APR-2001;
FEATURES Location/Qualifiers
source
1..1998
/organism="unknown"

BASE COUNT 410 a 596 c 568 g 424 t
ORIGIN

Alignment Scores:

Pred. No.: 1.46e-63 Length: 1998
Score: 864.50 Matches: 165
Percent Similarity: 92.35% Conservative: 12
Best Local Similarity: 84.18% Mismatches: 3
Query Match: 85.09% Indels: 1
DB: 6 Gaps: 1

US-09-965-594-22 (1-197) x AR145266 (1-1998)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
|||||
Db 64 GGTTCGTGTTATTGTTGTAGTAATATTTATCTGTTAGTGTAGTATCATCGGCGCTAC 123
|||||
QY 22 AlaGlnGlnThrArgGlyGluGlnGlnThrGlnLysThrSerHisThrGlyArgAspLys 41
:::|||||
Db 124 TCCCAACAGACGCGGGGCGCTACTTGGTTCGCAAGATCACTAGCCTTACAGGCGGGGACAG 183
|||||
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
|||||
Db 184 AACAGGTCGAGGAGAGGTTTCAGTGGTTTCCACCGCAACACAACTCCTTCTGGCGACC 243
|||||
QY 62 SerIleAsnGlyValLeuThrPheValThrHisGlyAlaGlyThrArgThrIleAlaSer 81
:::|||||
Db 244 TGGGTCAACGCGGTGTTGGACCGTTTACCATGGTGTGGCTCAAAAGACCTTAGCGCGC 303
|||||
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101
|||||
Db 304 CCAAGGGGCGCAATCCACAGATGTACATAATGTGGACAGACCTCGTGGCGTGGCAG 363
|||||
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
|||||
Db 364 CGCGCCCGCGGCGCGTTCCTTGACACCATGTCACCTGTGGCAGCTCAGACCTTTACTTG 423
|||||
QY 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141
|||||
Db 424 GTACAGAGACATCTGACGTCTCCGTGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCG 483
|||||
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
|||||
Db 484 CTCTCCCGCAGGCGTGTCTCTACTTGAAGGGCTCTCGGGTGTCTCCACTGCTCTGCCCT 543
|||||
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
:::|||||
Db 544 TCGGGGCACGCTGTGGCATCTTCGGGCTGCGGTATGCACCGGGGGGTTCGGAAGGCG 603
|||||
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
|||||
Db 604 GTGGACTTGTGCCCGTAGAGTCCATGGAAACTACTATGCGGTCTCCG 651

RESULT 11
AR145261
LOCUS AR145261 1998 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 102 from patent US 6211338.
ACCESSION AR145261
VERSION AR145261.1 GI:15107128
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.

REFERENCE
Unclassified.
1 (bases 1 to 1998)
AUTHORS Malcolm,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.
TITLE Single-chain recombinant complexes of hepatitis C virus NS3
protease and NS4A cofactor peptide
JOURNAL Patent: US 6211338-A 102 03-APR-2001;
FEATURES Location/Qualifiers
source
1..1998
/organism="unknown"

BASE COUNT 409 a 597 c 567 g 425 t
ORIGIN

Alignment Scores:

Pred. No.: 1.77e-63 Length: 1998
 Score: 863.50 Matches: 166
 Percent Similarity: 91.84% Conservative: 14
 Best Local Similarity: 84.69% Mismatches: 13
 Query Match: 84.99% Indels: 3
 DB: 6 Gaps: 1

US-09-965-594-22 (1-197) x ARI45261 (1-1998)

QY 5 GlySerValValIleValGluValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
 DB 64 GGTCTGTTGTTATTTGTTGTTAGATATTTTATCTGTTAGTGTAGTATCATCGGCTAC 123
 QY 22 AlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAspLys 41
 DB 124 TCCCAACAGACCGCGGCTACTTGGTTGTCATCATCTACAGCTTACAGCGCGGACAAG 183
 QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
 DB 184 AACAGGTGAGGAGAGGTTTCAGGTGTTTCCACCGCAACACAAATCTTCTGGCGACC 243
 QY 62 SerIleAsnGlyValLeuThrThrValThrHisGlyAlaGlyThrArgThrIleAlaSer 81
 DB 244 TGGCTCAACGGCGTGTGTGACCGCTTACCATGGTGTGCTGCCTCAAGACCTTAGCGCGC 303
 QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrP6ln 101
 DB 304 CCAAGGGGCAATCACCCAGATGTACACTAATGTGGACAGGACCTCGTCGGCTGGCAG 363
 QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
 DB 364 GCGCCCCCGGGCGGCTTCTTGACACCATGTGCACCTGTGGCAGCTCAGACCTTTACTTG 423
 QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
 DB 424 GTCAGGACATGCTGACGTCATTCGCGTGGCGGGCGGCGACAGTAGGGGAGCGCTG 483
 QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
 DB 484 CTCCTCCCGCCAGCGCTGCTCTACTTGAAGGGCTCTTCGGGTGGTCCACTGCTGCGCT 543
 QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
 DB 544 TCGGGGACCGTGTGGGATCTTCGGGCTCCGATGTCACCCGGGGGTTGCGAAGCG 603
 QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
 DB 604 GTGGACTTTGTGCGGTAGATCCATGGAACACTACTATGCGGTCTCCG 651

RESULT 12
 ARI45269 2016 bp DNA linear PAT 08-AUG-2001
 LOCUS
 DEFINITION Sequence 110 from patent US 6211338.
 ACCESSION ARI45269
 VERSION ARI45269.1 GI:15107136
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.

REFERENCE
 1 (bases 1 to 2016)
 AUTHORS Malcolm,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.
 TITLE Single-chain recombinant complexes of hepatitis C virus NS3
 protease and NS4 cofactor peptide
 JOURNAL Patent: US 6211338-A 110 03-APR-2001;
 FEATURES Location/Qualifiers
 source 1..2016
 BASE COUNT 412 a 603 c 570 g 431 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1.79e-63 Length: 2016
 Score: 863.50 Matches: 166

Percent Similarity: 91.84% Conservative: 14
 Best Local Similarity: 84.69% Mismatches: 13
 Query Match: 84.99% Indels: 3
 DB: 6 Gaps: 1

US-09-965-594-22 (1-197) x ARI45269 (1-2016)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
 DB 82 GGTCTGTTGTTATTTGTTGTTAGATATTTTATCTGTTAGTGTAGTATCATCGGCTAC 141
 QY 22 AlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAspLys 41
 DB 142 TCCCAACAGACCGGGGCTACTTGGTTGTCATCATCTACAGCTTACAGCGCGGACAAG 201
 QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
 DB 202 AACAGGTGAGGAGAGGTTTCAGGTGTTTCCACCGCAACAAATCTTCTGGCGACC 261
 QY 62 SerIleAsnGlyValLeuThrThrValThrHisGlyAlaGlyThrArgThrIleAlaSer 81
 DB 262 TGGCTCAACGGCGTGTGTGACCGCTTACCATGGTGTGCTGCCTCAAGACCTTAGCGCGC 321
 QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrP6ln 101
 DB 322 CCAAGGGGCAATCACCCAGATGTACACTAATGTGGACAGGACCTCGTCGGCTGGCAG 381
 QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
 DB 382 GCGCCCCCGGGCGGCTTCTTGACACCATGTGCACCTGTGGCAGCTCAGACCTTTACTTG 441
 QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
 DB 442 GTCAGGACATGCTGACGTCATTCGCGTGGCGGGCGGCGACAGTAGGGGAGCGCTG 501
 QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
 DB 502 CTCCTCCCGCCAGCGCTGCTCTACTTGAAGGGCTCTTTCGGGTGGTCCACTGCTGCGCT 561
 QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
 DB 562 TCGGGGACCGCTGTGGCATCTTCGGGCTGCGGTATGCACCCGGGGGTTGCGAAGCG 621
 QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
 DB 622 GTGGACTTTGTGCGCGTAGATCCATGGAACACTACTATGCGGTCTCCG 669

RESULT 13
 AF268278 12734 bp RNA linear VRL 12-JUL-2000
 LOCUS
 DEFINITION Pestivirus type 1, complete genome.
 ACCESSION AF268278
 VERSION AF268278.1 GI:9049956
 KEYWORDS
 SOURCE Pestivirus type 1
 ORGANISM Pestivirus type 1
 Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

REFERENCE
 1 (bases 1 to 12734)
 AUTHORS Lai,V.C., Zhong,W., Skelton,A., Ingravallo,P., Vassiliev,V.,
 Denis,R.O., Hong,Z. and Lau,J.Y.
 TITLE Generation and characterization of a hepatitis C virus NS3
 protease-dependent bovine viral diarrhea virus
 JOURNAL J. Virol. 74 (14), 6339-6347 (2000)
 MEDLINE 20323484
 PUBMED 10864644
 REFERENCE
 2 (bases 1 to 12734)
 AUTHORS Lai,V.C.H. and Hong,Z.
 TITLE Direct Submission
 JOURNAL Submitted (16-MAY-2000) Antiviral Therapy, Schering-Plough Research
 Institute, 2015 Galloping Hill Road, Kenilworth, NJ 07033-0539, USA
 FEATURES
 source 1..12734

/organism="Pestivirus type 1"

/mol_type="genomic RNA"

/db_xref="taxon:11099"

1..385

386..12508

/codon_start=1

/product="polyprotein"

/protein_id="AAF82566.1"

/db_xref="GI:9049957"

/translation="MELTNGSGSVVIVGRIVLSGSGSITACAOQTRGLLCKLITSL
TGRDKQVEGOIVSTAQOTELATCINGVCHTVYHAGTRIASPKGVIOYINVD
QDLVNPAPGCSRLIPICSGSDLYLTHANVIVRRRGSNGLSPRLSTLKG
SSGGLPDLCPAGHAGVLPRAACVTRGAKAVDFIPVENLETTIRSGADTEVCCSM
SYDTEEGATKTKQKPDRLGRKMKIVPKSEKSKTPDPDATVYGVYQVRKK
GKTSKNTDGLYHNKKNPOESRKKLEKALAIIVLFOVTGENTIQNLQDNG
TEGIRAMFORGVNRLHGLPEKICTGVPSHLATDIELKTHGMMADEKNTYCCR
LORHEWNGHCWNYNIEPLVMNRTQANLTGEGOPRECAVTCRDVDSLNVYDGL
RDSPTPLTCCCKKNFSFAGILMRGCPNFEIAASDVLPRHEHRISHPQDTLLVDGL
TNSLEGARQKATKLTWLGKQGLIGLKKLENKSKTWFGAYASPYCDVDKIGIYIWT
KNCPTACLPKATKIIVGPGFDTNAEDGKILHEMGHLSVLLGLSLVLSDDFAPEYASV
MYLILHFSIQSHVDVMDCKDTQNLNTVELTITADVIPGSMNGLKWCVRPNMWPEYET
TVVLAFEVSOVKVLRALRDLTRIWNAAITTAFLICLVKIVRGOMVGLINWLLIT
VGVOHLDCXPEFSYATAKDERIGOLAGELTTTWKEYSPCKLEDTMTWAKCEDKLM
YLQCRTRERLAILTHRALPTSVVPKCLFDGKQEDVVMNDNFEFGCLPCDAPIV
RGFNTLLGPFAPQWCPIGWTVGTSYCSFNMDLATIVRTYRKSFPFPHRQGCIT
QKNGEDLHNCILGNNMTCVPGDOLLYKGSIKESCGYQFKESGLPHYPIGKCL
ENETGYLVDTSNRSNGVAIVPQGLKCKIGKTVQVIAMDTKLGPMPCRYEIKLS
EGVEKTAFTNTKTKYFEPDSYFOQYMLKGEYOTWFDLEVTDRHRYFAESI
LVVVALLGGRYVLLVTVMLSEOKALGIOYGSVEVMGMLTHNNIEVTVYVFL
LYLLREESVKKVLLYHILVHPKSVIILIMIGDVVKADSGGOEVLKIDLCFT
TVYLVIGLLIARSDPTVPLVIMAAARVTELTQPGVDIAVAVMIIILLAVSYVD
YFRKMWQOGLISGVFLIRSLIYIGRIEMPEVTIPNWRPLTLLILYISTTVTR
WKYDVAGLLQCVPIPLLVTLLWADFLILIPYLYELKLYLKYTRVDIERSWLG
IDYTRVDSYDVSEGVYFPFROKAGNFSILLPLIKATLISCVSQWOLITMSY
LITDFMYHRKYIEISGTNIIISRLVAALIELNMBEESKGLKLYLLSGRLN
LITKHVRNFTSVYGEVYGMPIIITIKATLSKSRHCLICITCEGRWKWGC
PKGRHCKPTCCMSLADPERHYKRIEFGNFCMCRSCOCGRHFEWDEPKSAR
YCAECNRLHAPEGDFAESSMLGLKITPALMDGVYDITWAGCQRGVISPDTHRL
PCHISFGSRPPEQYNGVQYTAGOLFRLNPLVATIKVLMNNGEENIEHL
GHLRGPAVCKITHEKCHINDLITAFGIMPRGTTTPRAPREPTSLLYRROGLE
TGWATHQGSISSYDHTAGKDLVDSMGRTRVQCSNNRLTDEYGVKTVDSGCPD
GARCYLVNPAVNISSGKAVHLOKIGBCTVTSAGTAPFDFLANLKGWSGLPFE
ASSGRVGVKVGKNSKPKTUMSGIOTVSKNTADLTENVKKTSMNRCDFKQITLA
TGAGKTTELKAVIEIGRHKRVLVLIPIRAAESVYQYARLKHPSISFNLRIGDNKE
GDMAITASYSTGTCOMPQKRNAAVEYSYIFLDEHCAETPEQALIGKTHRFSES
IRVAMTATPAGSYTCQKHPITEEFAPEVMKGEDLSQFLDINGLKIPVDMKGM
LVFVTRMAVEAKKLKAKYNSGYISGEDPANRVVTSQSPYIVATNAIESGV
LPDLDTVIDTCLACERVRVSSKIPFIVTGLKRMVTVGEQORGRVGRVPGYR
SQETATSGDYHDLQAOARGTIDGINTKSPRENNYDWSLYEEDSLITOLELNN
LLISEDLPAVKNIARTDHPETIOLAYNSYEQVPLPPKIRNGEVTDYENYFLN
ARKLGEDVPYIVATEDEDLAVLLGLDWDPDGNOQVETGKALKQVTLGLSSENALL
VALFGYVQALSKRHPMIDITITIEDQLEDTHLOAPNAIKTDGTETELKELAS
GDVEKIMGASDAAGGELEFVSOAEKIKTAPLFKENAAGYQKFIIDSLIENKEE
IIRYGLDTHALYKISAAIRHETAFATLVKLWLPAGGESVDHYKQAADVLYVYV
MKNPSFGDSETOEGRFVASLFIASALATYTYKTNHNSKVVEPALAYLPYATSA
LKMFTPRLESVLIITIKYLSIRKSGDLGTLGISAAMEILSQNPVSVGISVM
LGWGAIAHNAIESSEKRTLLKMYVKNFLDQAAATDELVKENKPIIMALEFAVOTI
GNPLRLYHLGYVYKWEAKELSERTAGRNLTILMEFAPELLGMDSGKIRNLNGN
YILDLIYGLHQKIRLKKMVLGWAFFPCSDMTPSDERIRLPTDNLIRIKPVAKLEGOVE
EMAFKNVNGGKLTKVEESGFLCNRNPGFVNVRYTVYDDNLRIRIKPVAKLEGOVE
HYKGVATKIDYSGKMLLATDKVEHGVITRLAKRYTGVGPNGLBEKEIPTATVTWL
ERDCATITTKTVQFLKMKKCAFTYDITISNIRLIELVHNNLEBEKEIPTATVTWL
AYTEVNEDEVCTIKPVLGERVLPDPVVDINLOPEVOVDTSFVGITITIGRETMTGVT
VLEKVEPADSNQNSYKIGLEDGNYPGPGIQTHTLEIHNDRARFPIMLSGRNSIS
NRAKTARNILYTDNREIDLMAAGMLVVALROVDPELSEWDFKTFDLDREALE
ALSUGPKPKQVKEAYNIEQKQVEIENWFASDDPVFEALKNDRKYLIVGDYGE
VKDQAKALGATDQTRIIKEVSGTAYMKLSNFWASQNSQMSLTPLEILLRCPAT
KSNKGHMASYQLAOGNWEPLGCGVHLGTPARRVKIHPYEAVLKLKLDFTBEBEKPR
VKDTPVIRHNKWLKLTIRFOGNLNTKMLNPKGLSQLDREGRKNYINHQIGTINSS
AGIRLEKLPYRAQDTKTTHEAIRDKIKSENQNPENLHNKLEIFHTIAQTLKHT
YGEVTEQLAGENRKGAAQFLEKKNIAGEVDEKHLVEOLVRDKLAKRYETFAI
PKNEKROVDQMDQALVVEKRPRIQYPEAKTRLAITKYNNVNRKQOPVPIPYEBC
TPLNFIDKVRKENDSFNEPVAVSFTKAMDQVTSKDLQICEIGIYKYYKEMHNF

DTITDHMEVPIVITADGEYVIRNQQRSGQPDYSAGNSMLNVLNMTMYAFCESTGVYPK
SFNEVARIHVCDDGDFLITBKGLCKPANKQMOLHEAGKPKQITTECKMKVAFREFD
IEFCSHTPVPVVRWSDNTSSHWAGRDATVILSKMATRLDSSGERGTTAYEKAVAFSL
MYSNPLVRICLLVLSQOPETDPSKHATYYKGDPIGAYKDVIGNLSLSEKRTGEK
LANLKSUSLIGWIKHTSKRIQDCVAIGKEGNMLYNADRLISSKTHGLYIPDGF
TLQKHYYEQLOLRTETNPVNGVTERYKLGPIVNLRLRLKILLMTAVGVSS
12509..12734

BASE COUNT 4030 a 2608 c 3293 g 2802 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 12734
Score: 862.50
Percent Similarity: 91.28%
Best Local Similarity: 88.72%
Query Match: 84.89%
DB: 14
Length: 12734
Matches: 173
Conservative: 5
Mismatches: 14
Indels: 3
Gaps: 1

US-09-965-594-22 (1-197) x AF268278 (1-12734)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
|||||
Db 413 GGTAGTGTGTATTGTTGGTAGAATGTTTATCTGTTAGTATATCAGCGCGTC 472
QY 22 AlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAspLys 41
|||||
Db 473 GCCCAGCAGCAGCAGCGCTCTAGGTGTAAAGATCACCAGTCTGACTGCGCGGACAAA 532
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
|||||
Db 533 AACCAAGTGGAGGGTGGAGTCCAGATCGTCACTGTACCCAAACCTTCTCGGCAACG 592
QY 62 SerIleAsnGlyValLeuThrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
|||||
Db 593 TGCATCAATGGGTATGCTGGACTGCTACACAGGGCGCGACAGGACCATCGCATCA 652
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101
|||||
Db 653 CCCAAGGTCTCTCATCATCATGATATACCAATGTGGACCAAGACCTTGTGGGCTGCCCC 712
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
|||||
Db 713 GCTCTCAAGGTTCCCGTCTATGACACCCCTGACCTCGCGCTCTCGGACCTTACTGT 772
QY 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141
|||||
Db 773 GTTAGAGGACAGCCCAAGCTATTCCCTGCGCGCGGCGAGGTAGTAGCGGGTAGCGTG 832
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
|||||
Db 833 CTTTCGCCCGCCCATTTCTTACCTAAAGGCTCTCTGGGGGTCCGCTGTGTGTGCCCC 892
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
|||||
Db 893 CGGGGACAGCGCGTGGGCTATTACAGGCGCGGTGTACACCGCTGGAGTGGCCAAAGCG 952
QY 182 ValAspPheIleProValIleSerLeuGluThrThrMetArgSer 196
|||||
Db 953 GTTGACTTTATCCCTGTGGGAACCTAGAGACAAACACGAGATCC 997

RESULT 14

LOCUS

ARI45257 651 bp DNA linear PAT 08-AUG-2001
Sequence 98 from patent US 6211338.

DEFINITION

ARI45257

ACCESSION

ARI45257.1 GI:15107124

VERSION

KEYWORDS

UNKNOWN.

SOURCE

ORGANISM

UNKNOWN.

REFERENCE

1 (bases 1 to 651)

AUTHORS

Malcolm, B.A., Taremi, S., Shane, J., Weber, P.C. and Yao, N.

TITLE

Single-chain recombinant complexes of hepatitis C virus NS3

protease and NS4A cofactor peptide
Patent: US 6211338-A 98 03-APR-2001;

JOURNAL

FEATURES
source Location/Qualifiers

BASE COUNT 119 a 188 c 199 g 145 t
ORIGIN
1. .651
/organism="unknown"

Alignment Scores:

Pred. No.: 7.16e-64 Length: 651
Score: 861.50 Matches: 165
Percent Similarity: 92.31% Conservatives: 15
Best Local Similarity: 84.62% Mismatches: 12
Query Match: 84.79% Indels: 3
DB: 6 Gaps: 1

US-09-965-594-22 (1-197) x AR145257 (1-651)

```
Qy 5 GlySerValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
|||||
Db 64 GGTTCGTGTTATGTTGGTAGAATATTTATCTGGTAGTGTAGTATCATCGGCGCTAC 123
|||||
Qy 22 AlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAspLys 41
|||||
Db 124 TCCCAACAGACGGCGGCTACTTGGTTCATCAGACTAGCCTTACAGCGCGGACAAG 183
|||||
Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
|||||
Db 184 AACAGGTCAGGGAGAGGTTTCAGTGGTTTCCACCGCAACACAATCCTTCCTGGCGACC 243
|||||
Qy 62 SerIleAsnGlyValLeuTyrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
|||||
Db 244 TGGGTCAACGGCGTGTGGACCGTTTACCATGGTGGCTCAAGACCTTAGCGCGC 303
|||||
Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101
|||||
Db 304 CCAAGGGGCAATCACCAGATGTACACTAATGTGGACGAGGACCTCGTGGCTGGCAG 363
|||||
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
|||||
Db 364 CGCGCCCGCGGCGCGTCTCTTGACACCATGCACCTGTGGCAGCTCAGACCTTTACTTG 423
|||||
Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
|||||
Db 424 GTCACGAGACATGCTGACGTCATTCCGGTGGCGGCGGCGGCGGCGGAGGCTG 483
|||||
Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
|||||
Db 484 CTCCTCCCGGCGCTGCTCTCTACTTGAAGGGCTCTGCTGTGTCCACTGCTCGCCT 543
|||||
Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
|||||
Db 544 TCGGGGACCGTGTGGCATCTCCGGGCTCCCGTATGCACCGCGGGGTTGCCAAGCGC 603
|||||
Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196
|||||
Db 604 GTGGACTTTGGCCCGTAGAGTCCATGGAACACTACTATGCGGTCT 648
```

RESULT 15

AR145252

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AR145252 651 bp DNA linear PAT 08-AUG-2001

Sequence 93 from patent US 6211338.

AR145252.1 GI:15107119

Unknown.

Unclassified.

1 (bases 1 to 651)

Malcolm, B.A., Taremi, S., Shane, Weber, P.C. and Yao, N.

Single-chain recombinant complexes of hepatitis C virus NS3

protease and NS4A cofactor peptide

Patent: US 6211338-A 93 03-APR-2001;

Search completed: August 31, 2003, 00:46:41

Job time : 2569.57 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 30, 2003, 19:13:57 ; Search time 182.939 Seconds
(Without alignments)
2906.924 Million cell updates/sec

Title: US-09-965-594-22

Perfect score: 1016

Sequence: 1 MKKKGSVIVGRINLSGDTA.....VAKAVDFIPVESLETTMRSP 197

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xip
-Q=/cgn2_1/USPTO.spool/US09965594/runat_29082003_151918_28302/app_query.fasta_1.2872
-DB=N.Geneseq.19Jun03 -OPMT=fastap -SUFFIX=ing -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human0.cdi
-LIST=45 -DOALIGN=200 -THK=MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09965594 -CGN_1_1_1412 -runat_29082003_151918_28302 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP-6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N.Geneseq.19Jun03.*

- 1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
- 25: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	1016	100.0	594	21	AAA73334
2	1010	99.4	594	21	AAA73333
3	995	97.9	594	21	AAA73332
4	980	96.5	594	21	AAA73331
5	963	94.8	594	21	AAA73330
6	936	92.1	594	21	AAA73329
7	929	91.4	588	21	AAA73328
8	902	88.8	588	21	AAA73327
9	882.5	86.9	12734	24	ABA95615
10	875.5	86.2	1998	20	AAH80355
11	872.5	85.9	1998	20	AAH80359
12	871.5	85.8	1998	20	AAH80354
13	868.5	85.5	651	20	AAH80345
14	868.5	85.5	1998	20	AAH80358
15	867.5	85.4	1998	20	AAH80353
16	865.5	85.2	651	20	AAH80349
17	864.5	85.1	612	25	ABX15706
18	864.5	85.1	651	20	AAH80344
19	864.5	85.1	1998	20	AAH80357
20	863.5	85.0	1998	20	AAH80352
21	863.5	85.0	2013	20	AAH80360
22	861.5	84.8	651	20	AAH80348
23	860.5	84.7	651	20	AAH80343
24	860.5	84.7	1998	20	AAH80356
25	860.5	84.7	2016	20	AAH80361
26	860	84.6	648	20	AAH80365
27	858	84.4	648	20	AAH80363
28	857.5	84.4	650	20	AAH80347
29	857.5	84.4	651	20	AAH80351
30	856.5	84.3	651	20	AAH80342
31	854	84.1	648	20	AAH80362
32	853.5	84.0	650	20	AAH80346
33	853.5	84.0	651	20	AAH80350
34	851	83.8	8145	20	AAH23259
35	849	83.6	1933	20	AAH23258
36	848.5	83.5	9646	19	AAV59361
37	848.5	83.5	9646	24	ABR87285
38	848.5	83.5	12980	19	AAV59364
39	848.5	83.5	12980	24	ABR87286
40	848.5	83.5	16622	21	AAZ36212
41	844.5	83.1	5300	10	AAH92097
42	844.5	83.1	5360	10	AAH90327
43	844.5	83.1	6905	10	AAH92103
44	844.5	83.1	7310	10	AAH92106
45	844.5	83.1	7310	10	AAH90336

ALIGNMENTS

RESULT 1	
AAA73334	
ID	AAA73334 standard; DNA; 594 BP.
XX	AAA73334;
XX	
XX	19-DEC-2000 (first entry)
XX	
DE	Hepatitis C virus NS4A-NS3 fusion protease coding sequence #7.
XX	
KW	Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW	liver failure; liver cancer; mutant; muteln; ds.
XX	
OS	Hepatitis C virus.
OS	Synthetic.
XX	
XX	
XX	Key
XX	Location/Qualifiers

```

FT CDS 1..594
FT /*tag- a
FT /product= "NS4A-NS3 fusion protein #7"
XX WO200040707-A1.
XX 13-JUL-2000.
XX 06-JAN-2000; 2000WO-US00345.
XX 08-JAN-1999; 99US-0115271.
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX WPI; 2000-465976/40.
XX P-PSDB; AAB15225.
XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
XX substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
XX amino acid, useful for screening inhibitors that may treat hepatitis C
XX
XX Claim 26; Fig 17; 66pp; English.
XX The present sequence is the coding sequence for a mutated version of a
XX fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
XX protease enzymes. These proteins are both essential for the replication
XX of the virus, acting to cleave its replicative proteins from the
XX polyprotein produced from the HCV genome. Inhibitors of the two proteins
XX should be effective as antiviral treatments of HCV infection. This is
XX useful as HCV can lead to chronic liver disease such as cirrhosis, liver
XX failure and liver cancer. The present invention concerns a number of NS3
XX mutants and NS3-NS4A fusion proteins which can be used to identify
XX inhibitors of this type, as well as enabling structural studies of the
XX protease and protease:inhibitor complexes. The protein produced from this
XX sequence contains the alpha-helix0-7 variant.
XX
XX Sequence 594 BP; 105 A; 192 C; 151 G; 146 T; 0 other;
XX
Alignment Scores:
Pred. No.: 3,95e-86 Length: 594
Score: 1016.00 Matches: 197
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 21 Gaps: 0
US-09-965-594-22 (1-197) x AAA73334 (1-594)
Qy 1 MetLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
Dy 1 ATGAAAAAAGGATCCGTTGTTATCGTCGCGGTATCAACCTGTCGGTGACACCGCT 60
Qy 21 TyrAlaGlnGlnThrArgGlyGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 40
Dy 61 TAGCCTCAGCAGACTCGAGTGAGCAGGGTACCAGAACACCTCCACACCGCTGCTGAC 120
Qy 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
Dy 121 AAAAACAGGTTCAAGGTGAAGTTCAGATCGTTTCCACCGCTACCCAGACCTTCCTGGCT 180
Qy 61 ThrSerIleAsnGlyValLeuThrThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
Dy 181 ACCTCCATCAACGGTGTCTGTGGACGGTTTACACGGTGTGTACCGGTACCATCGCT 240
Qy 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr 100
Dy 241 TCCCCGAAAGTCCGGTATCCCAAGATGTACACCAAGCTGTACAAAGACCTGGTGTGG 300
Qy 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
Dy
301 CAGGCTCCGAGGGTTCCTGCTGACCCGCTGACCTGCGGTTCCTCCGACCTGTAC 360
121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
361 CTGGTTACCGCTACGCTGACGTTATCCCGGTTCGTCGTGCTGCTGCTGCTGCTGCT 420
141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
421 CTGCTGTCCCGCGTCCGATCTCTACCTGAAGGTTCTCCGCTGCTGCTGCTGCTGCT 480
161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180
481 CCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
541 GCTGTGTGACTTCATCCCGTTGAATCCCTGGAAACCAACCATGCGTTCCTCCCG 591
RESULT 2
AAA73333
ID AAA73333 standard; DNA; 594 BP.
XX AAA73333;
XX
XX 19-DEC-2000 (first entry)
XX Hepatitis C virus NS4A-NS3 fusion protease coding sequence #6.
XX
XX Hepatitis; NS3 protease; viral replication; chronic liver disease;
XX liver failure; liver cancer; mutant; mutein; ds.
XX
XX Hepatitis C virus.
XX Synthetic.
XX
XX Key Location/Qualifiers
XX CDS 1..594
XX FT /tag- a
XX FT /product= "NS4A-NS3 fusion protein #6"
XX
XX WO200040707-A1.
XX
XX 13-JUL-2000.
XX 06-JAN-2000; 2000WO-US00345.
XX 08-JAN-1999; 99US-0115271.
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX WPI; 2000-465976/40.
XX P-PSDB; AAB15224.
XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
XX substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
XX amino acid, useful for screening inhibitors that may treat hepatitis C
XX
XX Claim 26; Fig 16; 66pp; English.
XX The present sequence is the coding sequence for a mutated version of a
XX fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
XX protease enzymes. These proteins are both essential for the replication
XX of the virus, acting to cleave its replicative proteins from the
XX polyprotein produced from the HCV genome. Inhibitors of the two proteins
XX should be effective as antiviral treatments of HCV infection. This is
XX useful as HCV can lead to chronic liver disease such as cirrhosis, liver
XX failure and liver cancer. The present invention concerns a number of NS3
XX mutants and NS3-NS4A fusion proteins which can be used to identify
XX inhibitors of this type, as well as enabling structural studies of the
XX protease and protease:inhibitor complexes. The protein produced from this
XX sequence contains the alpha-helix0-7 variant.
XX
XX Sequence 594 BP; 105 A; 192 C; 151 G; 146 T; 0 other;
XX
Alignment Scores:
Pred. No.: 3,95e-86 Length: 594
Score: 1016.00 Matches: 197
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 21 Gaps: 0
US-09-965-594-22 (1-197) x AAA73334 (1-594)
Qy 1 MetLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
Dy 1 ATGAAAAAAGGATCCGTTGTTATCGTCGCGGTATCAACCTGTCGGTGACACCGCT 60
Qy 21 TyrAlaGlnGlnThrArgGlyGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 40
Dy 61 TAGCCTCAGCAGACTCGAGTGAGCAGGGTACCAGAACACCTCCACACCGCTGCTGAC 120
Qy 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
Dy 121 AAAAACAGGTTCAAGGTGAAGTTCAGATCGTTTCCACCGCTACCCAGACCTTCCTGGCT 180
Qy 61 ThrSerIleAsnGlyValLeuThrThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
Dy 181 ACCTCCATCAACGGTGTCTGTGGACGGTTTACACGGTGTGTACCGGTACCATCGCT 240
Qy 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr 100
Dy 241 TCCCCGAAAGTCCGGTATCCCAAGATGTACACCAAGCTGTACAAAGACCTGGTGTGG 300
Qy 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
Dy

```

XX SQ Sequence 594 BP; 104 A; 191 C; 152 G; 147 T; 0 other;

Alignment Scores:
 Pred. No.: 1,44e-85 Length: 594
 Score: 1010.00 Matches: 196
 Percent Similarity: 99.49% Conservatives: 0
 Best Local Similarity: 99.49% Mismatches: 1
 Query Match: 99.41% Indels: 0
 DB: 21 Gaps: 0

US-09-965-594-22 (1-197) x AAA73332 (1-594)

QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
 DB 1 ATGAAAAAAGGATCGGTGTTATCGTCCGCCGTATCAACCTGTCGGGTGACACCGCT 60
 QY 21 TyrAlaGlnGlnThrArgGlyClnGlnGlnThrSerHisThrGlyArgAsp 40
 DB 61 TACGCTCAGCAGCTCGAGGTGAGCAGGTTGCCAGAGACCTCCACACCGGTGCTGAC 120
 QY 41 LysAsnGlnValGluGlyValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
 DB 121 AAAACACAGGTGAAGTGAAGTTCAGATCGTTCCACCGGTACCCAGACCTTCCTGGCT 180
 QY 61 ThrSerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
 DB 181 ACCTCATCAACGGGTGTTCTGTGGACCGTTTACCACGGTGTGCTACCCGTACCATCGCT 240
 QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp 100
 DB 241 TCCCGAAGGTGCCGTGTACCCAGATGTACCAACGTTGACAAAGACCTGTTGGTTGG 300
 QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
 DB 301 CAGGCTCCGACGGTTCGGTTCCTCGACCCGTCACCTGCGGTTCCTCCGACCTGTAC 360
 QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
 DB 361 CTGGTTACCCGTCACCGCTGACGTTATCCCGGTGCTGCTGCTGCTGCTGCTGCTGCT 420
 QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
 DB 421 CTGCTGTCCCGCTCCGATCTCTACCTGGAAGGTTCCTCCGGTGGTCCGCTGCTGTC 480
 QY 161 ProAlaGlyHisAlaValClyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180
 DB 481 CCGGCTGTGTCACCGCTGTTGGTATCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
 QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
 DB 541 GCTGTTGACTTCATCCCGGTGAATCCCTGGAAACACCATGCTGCTCCCG 591

RESULT 3

AAA73332

ID AAA73332 standard; DNA; 594 BP.

XX AC AAA73332;

XX DT 19-DEC-2000 (first entry)

XX DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #5.

XX KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
 liver failure; liver cancer; mutant; mutein; ds.

XX OS Hepatitis C virus.

OS Synthetic.

XX FH Key Location/Qualifiers

XX FT CDS 1..594

FT /tag= a

FT /product= *NS4A-NS3 fusion protein #5*

XX FN WO2000040707-A1.
 XX PD 13-JUL-2000.
 XX PF 06-JAN-2000; 2000WO-US00345.
 XX PR 08-JAN-1999; 99US-0115271.
 XX PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
 XX WPI; 2000-465976/40.
 DR P-PSDB; AAB15223.
 XX PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 .
 XX Claim 26; Fig 15; 66pp; English.
 CC The present sequence is the coding sequence for a mutated version of a
 CC fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
 CC protease enzymes. These proteins are both essential for the replication
 CC of the virus, acting to cleave its replicative proteins from the
 CC polypeptide produced from the HCV genome. Inhibitors of the two proteins
 CC should be effective as antiviral treatments of HCV infection. This is
 CC useful as HCV can lead to chronic liver disease such as cirrhosis, liver
 CC failure and liver cancer. The present invention concerns a number of NS3
 CC mutants and NS3-NS4A fusion proteins which can be used to identify
 CC inhibitors of this type, as well as enabling structural studies of the
 CC protease and protease-inhibitor complexes. The protein produced from this
 CC sequence contains the alpha-helix0-1 variant.
 SQ Sequence 594 BP; 105 A; 189 C; 153 G; 147 T; 0 other;

Alignment Scores:

Pred. No.: 3,64e-84 Length: 594
 Score: 995.00 Matches: 193
 Percent Similarity: 98.98% Conservatives: 2
 Best Local Similarity: 97.97% Mismatches: 2
 Query Match: 97.93% Indels: 0
 DB: 21 Gaps: 0

US-09-965-594-22 (1-197) x AAA73332 (1-594)

QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
 DB 1 ATGAAAAAAGGATCGGTGTTATCGTCCGCCGTATCAACCTGTCGGGTGACACCGCT 60
 QY 21 TyrAlaGlnGlnThrArgGlyClnGlnGlnThrSerHisThrGlyArgAsp 40
 DB 61 TACGCTCAGCAGCTCGAGGTGAGGAGGTGCCAAGAACTCCACAGACCGGTGCTGAC 120
 QY 41 LysAsnGlnValGluGlyValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
 DB 121 AAAACACAGGTGAAGTGAAGTTCAGATCGTTTCCACCGGTGCTGCTGCTGCTGCTGCT 180
 QY 61 ThrSerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
 DB 181 ACCTCATCAACGGGTGTTCTGTGGACCGTTTACCACCGGTGCTGCTGCTGCTGCTGCT 240
 QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp 100
 DB 241 TCCCGAAGGTGCCGTGTACCCAGATGTACCAACGTTGACAAAGACCTGTTGGTTGG 300
 QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
 DB 301 CAGGCTCCGACGGTTCGGTTCCTGACCCGTCACCTGCGGTTCCTCCGACCTGTAC 360
 QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140

```

|||||
361 CTGTTACCGTCACGCTGACGTTATCCGGTTCGTCGTGGTGACTCCCGTGGTCC 420
|||||
141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerClyClyProLeuLeuCys 160
|||||
421 CTGCTGTCCCGGTCGGATCTCCTACCGAAGGTCTCCGGTGGTCCGCTGTGTGC 480
|||||
161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180
|||||
481 CCGGCTGGTCACGCTGTGGTATCTTCCTCCGTCGTGTGTTCCACCGTGGTGTCTAA 540
|||||
181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
541 GCTGTTGACTTCATCCGGTGAATCCCTGGAAACCAACCATGCGTTCCCGG 591

RESULT 4
AAA73331
ID AAA73331 standard; DNA; 594 BP.
XX
AC AAA73331:
XX
DT 19-DEC-2000 (first entry)
XX
DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #4.
XX
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW liver failure; liver cancer; mutant; mutein; ds.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1..594
FT /tag= a
FT /product= "NS4A-NS3 fusion protein #4"
XX
PN W0200040707-A1.
XX
PD 13-JUL-2000.
XX
PF 06-JAN-2000; 2000MO-US0345.
XX
PR 08-JAN-1999; 99US-0115271.
XX
PA (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX
DR WPI: 2000-465976/40.
DR P-PSDB; AAB15222.
XX
PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
PT amino acid, useful for screening inhibitors that may treat hepatitis C
XX
XX
PS Claim 26; Fig 14; 66pp; English.
XX
XX The present sequence is the coding sequence for a mutated version of a
XX fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
XX protease enzymes. These proteins are both essential for the replication
XX of the virus, acting to cleave its replicative proteins from the
XX polyprotein produced from the HCV genome. Inhibitors of the two proteins
XX should be effective as antiviral treatments of HCV infection. This is
XX useful as HCV can lead to chronic liver disease such as cirrhosis, liver
XX failure and liver cancer. The present invention concerns a number of NS3
XX mutants and NS3-NS4A fusion proteins which can be used to identify
XX inhibitors of this type, as well as enabling structural studies of the
XX protease and protease-inhibitor complexes. The protein produced from this
XX sequence contains the alpha-helix0-1 variant.
XX
SQ Sequence 594 BP; 105 A; 187 C; 155 G; 147 T; 0 other;

```

```

Alignment Scores:
Pred. No.: 9.23e-83 Length: 594
Score: 980.00 Matches: 190
Percent Similarity: 97.46% Conservative: 2
Best Local Similarity: 96.45% Mismatches: 5
Query Match: 96.46% Indels: 0
DB: 21 Gaps: 0

US-09-965-594-22 (1-197) x AAA73331 (1-594)
QY 1 MetLysLysLysGlySerValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
DB 1 ATGAAAAAAGGATCGGTTGTTATCGTCGGCGGTATCAACCTGCGGGTGACACGCT 60
QY 21 TyrAlaGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyValAsp 40
DB 61 TAGCTCAGCAGACTCGAGGTGAGAGGTTGCCAAGAAACCTCCACACCGGTCGTGAC 120
QY 41 LysAsnGlnValGluGlyValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
DB 121 AAAAACCAGGTGAAGTGAAGTTCAGATCGTTCCACCGCTACCCAGACCTTCTCGCT 180
QY 61 ThrSerIleAsnGlyValLeuThrThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
DB 181 ACCTGCATCAACGGTGTTCGTGGACCGTTTACCACCGTGTCTGCCGTACCCGTACCATCGCT 240
QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValClyTyr 100
DB 241 TCCCGAAGGTCGCGTTACCCAGATGTACACCAACGTTGACAAAGACCTGGTTGGTTGG 300
QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
DB 301 CAGGCTCGCAGGGTTCGCTGACCGCTGCACCTCGCGTTCCTCCGAGCTGTCTCCGAGCTGTAC 360
QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
DB 361 CTGTTATCCCGTCACGCTGACGTTATCCCGTTCGTGTCGTGGTGACTCCCGTGGTTCC 420
QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyProLeuLeuCys 160
DB 421 CTGCTGTCCCGCGTCGCGATCTCTACCTGAAAGGTTCCTCCGCTGGTCCGCTGCTGTC 480
QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180
DB 481 CCGGCTGGTCACGCTGTTGGTATCTTCGTCGTCTGCTGCTTTCACCCCGTGTGTGCTAAA 540
QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
DB 541 GCTGTTGACTTCATCCCGGTTGAATCCCTGGAAACCAACCATGCGTTCCCGG 591

RESULT 5
AAA73330
ID AAA73330 standard; DNA; 594 BP.
XX
AC AAA73330:
XX
DT 19-DEC-2000 (first entry)
XX
DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #3.
XX
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW liver failure; liver cancer; mutant; mutein; ds.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1..594
FT /tag= a
FT /product= "NS4A-NS3 fusion protein #3"
XX
PN W0200040707-A1.
XX

```

```

PD 13-JUL-2000.
XX
PF 06-JAN-2000; 2000MO-US00345.
XX
PR 08-JAN-1999; 99US-0115271.
XX
PA (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX
DR WPI; 2000-465976/40.
XX
DR P-PSDB; AAB15221.
XX
PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
PT amino acid, useful for screening inhibitors that may treat hepatitis C
PT
XX
XX
PS Claim 26; Fig 13; 66pp; English.
XX
CC The present sequence is the coding sequence for a mutated version of a
CC fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
CC protease enzymes. These proteins are both essential for the replication
CC of the virus, acting to cleave its replicative proteins from the
CC polyprotein produced from the HCV genome. Inhibitors of the two proteins
CC should be effective as antiviral treatments of HCV infection. This is
CC useful as HCV can lead to chronic liver disease such as cirrhosis, liver
CC failure and liver cancer. The present invention concerns a number of NS3
CC mutants and NS3-NS4A fusion proteins which can be used to identify
CC inhibitors of this type, as well as enabling structural studies of the
CC protease and protease-inhibitor complexes. The protein produced from this
CC sequence contains the alpha-helix0-1 variant.
XX
SQ Sequence 594 BP; 103 A; 186 C; 156 G; 149 T; 0 other;

Alignment Scores:
Pred. No.: 3,59e-81 Length: 594
Score: 963.00 Matches: 187
Percent Similarity: 95.94% Conservative: 2
Best Local Similarity: 94.92% Mismatches: 8
Query Match: 94.78% Indels: 0
DB: 21 Gaps: 0

US-09-965-594-22 (1-197) x AAA73330 (1-594)

Qy 1 MetLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
Db 1 ATCAAAAAAAGATCCGTTGTATCGTCGGCGGTATCAACCTGTCGGGTGCACACCGCT 60
Qy 21 TyrAlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAsp 40
Db 61 TAGCCTCAGCAGACTCGAGGTGAGGAGGGTTGCAAGAAACCTCCCGACGCGGTGAC 120
Qy 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
Db 121 AAAACCCAGGTGAAGGTGAAGTTCAGATCGTTCCACCGCTCTCAGACCTTCCTGGCT 180
Qy 61 ThrSerIleAsnGlyValLeuThrThrValThrHisGlyAlaGlyThrArgThrIleAla 80
Db 181 ACCTGCATCAACGGTGTTCCTGGACCGTTTACACGGGTCTGCTGCTACCGGTACCATCGCT 240
Qy 81 SerProLysGlyProValThrGlnMetThrThrAsnValAspLysAspLeuValGlyTyr 100
Db 241 TCCCCGAAAGTCCGGTTATCCAGATGTACACCAACGTTTACAAAGACCTGGTGGTTGG 300
Qy 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
Db 301 CCGGCTCCGAGGGTTCCTGCTCCCTGACCCCGGTGACCTGCGGTTCCTCCGACCTGTAC 360
Qy 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
Db 361 CTGGTTACCCGTCACGCTGACGTTATCCCGGTTCCGTCGTCGTGTCGACATCCCGTGTTC 420

```

```

Qy 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
Db 421 CTGCTGTCCCGGTCGGATCTCCTACCTGAAGGTTCTCCGGTGGTGGTGGTGGTGC 480
Qy 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180
Db 481 CCGGCTGGTCACGCTGTGTGTATCTTCCGTGCTGCTGTGTGACCCGCTGGTGTGCTAAA 540
Qy 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
Db 541 GCTGTTGACTTCATCCGGTTGAATCCCTGGAAACACCATCGGTTCCCGC 591

RESULT 6
AAA73335
ID AAA73335 standard; DNA; 594 BP.
XX
AC AAA73335;
XX
DT 19-DEC-2000 (first entry)
XX
DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #8.
XX
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
XX liver failure; liver cancer; mutant; mutain; ds.
XX
OS Hepatitis C virus.
XX
FH Key Location/Qualifiers
FT CDS 1..594
FT /tag- a
FT /product- "NS4A-NS3 fusion protein #8"
XX
PN WO200040707-A1.
XX
PD 13-JUL-2000.
XX
PF 06-JAN-2000; 2000MO-US00345.
XX
PR 08-JAN-1999; 99US-0115271.
XX
PA (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX
DR WPI; 2000-465976/40.
XX
DR P-PSDB; AAB15226.
XX
PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
PT amino acid, useful for screening inhibitors that may treat hepatitis C
PT
XX
XX
PS Disclosure; Fig 18; 66pp; English.
XX
CC The present sequence is the coding sequence for a mutated version of a
CC fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
CC protease enzymes. These proteins are both essential for the replication
CC of the virus, acting to cleave its replicative proteins from the
CC polyprotein produced from the HCV genome. Inhibitors of the two proteins
CC should be effective as antiviral treatments of HCV infection. This is
CC useful as HCV can lead to chronic liver disease such as cirrhosis, liver
CC failure and liver cancer. The present invention concerns a number of NS3
CC mutants and NS3-NS4A fusion proteins which can be used to identify
CC inhibitors of this type, as well as enabling structural studies of the
CC protease and protease-inhibitor complexes. The protein produced from this
CC sequence contains the alpha-helix0 wild-type sequence.
XX
SQ Sequence 594 BP; 98 A; 189 C; 153 G; 154 T; 0 other;

Alignment Scores:
Pred. No.: 1.21e-78 Length: 594
Score: 936.00 Matches: 185

```

Percent Similarity:	93.91%	Conservative:	0
Best Local Similarity:	93.91%	Mismatches:	13
Query Match:	92.13%	Indels:	0
DB:	21	Gaps:	0

US-09-965-594-22 (1-197) x AAA73335 (1-594)

Qy	1	MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla	20
Db	1	ATGAAAAAAGAGATCGGTGTTATCGTGGCCGATCAACCTGCCGTGCACCGCT	60
Qy	21	TyrAlaGlnGlnThrArgGlyClnGlnGlyThrGlnLysThrSerHisThrGlyArgAsp	40
Db	61	TAGCTTCACGACACTCGAGGTCGCTGGTTGCATCATCACTCCCTGACCGGTCTGTGAC	120
Qy	41	LysAsnGlnValGlnGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla	60
Db	121	AAAAACCAAGTTGAAGGTCAAGTTCAGATCGTTCCACCGCTGCTCAGACCTTCCTGGCT	180
Qy	61	ThrSerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla	80
Db	181	ACCTGTCATCAACGGGTGTTTGTGTGGACCGTTTACCACGCTGCTGTGTACCCGTACCATCGCT	240
Qy	81	SerProLysGlyProValThrClnMetTyrThrAsnValAspLysAspLeuValGlyTrp	100
Db	241	TCCCCGAAAGGTCCGGTTATCCGATGTACACCAACGTTGACAAAGACCTGGTGTGGTGG	300
Qy	101	GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr	120
Db	301	CGGGCTCCGACGGTTCCTCCGTCACCCGCTGCACCTGCGGTCTCTCCGACCTGTAC	360
Qy	121	LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer	140
Db	361	CTGTTTACCCTGCACGCTGACGTTATCCCGGTTCTGCTGCTGTGTGTGTGTGTGTGT	420
Qy	141	LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys	160
Db	421	CTGCTGTCCCCGGTCCGATCTCTTACCTGAAGAGGTTCCTCCGGTGTGTGTGTGTGTGT	480
Qy	161	ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys	180
Db	481	CCGGCTGGTCACGCTGTGGTATCTCCGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	540
Qy	181	AlaValAspPheIleProValGlnSerLeuGluThrThrMetArgSerPro	197
Db	541	GCTGTGTGACCTTATCCCGGTTGAATCTCCCTGGAACCAACCATCGCTGCCCG	591

RESULT 7

AAA73329

ID AAA73329 standard; DNA; 588 BP.

AA
AC AAA73329:XX
XX
16755337

DT 19-DEC-2000 (first entry)

[illegible]

DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #2.

XX

KW Hepatitis; NS3 protease; viral replication; chronic liver disease;

KW liver failure; liver cancer; mutant; mutein; ds.

XX

OS Hepatitis C virus.

OS Synthetic.

XX

FH	Key	Location
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

CD5

```

/*tag= a
/*tag= a

```

References

XX
NW
2070000707-

PN 102000407-7
YY

13-III-2000

FD JUL 2000.
XX

06-JAN-2000: 2000WQ-UIS00345

7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844

XX	08-JAN-1999;	9905-0115271.	
XX	(BRIM)	BRISTOL-MYERS SQUIBB CO.	
XX	Wittekind M,	Weinheimer S, Zhang Y, Goldfarb V;	
XX	WPI;	2000-465976/40.	
XX	P-PSDB;	AAB15220.	
XX	Modified hepatitis C virus (HCV) NS3 protease comprising at least 1		
PT	substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic		
PT	amino acid, useful for screening inhibitors that may treat hepatitis C		
XX			
XX	Claim 26; Fig 12;	66pp; English.	
XX	The present sequence is the coding sequence for a mutated version of a		
CC	fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A		
CC	protease enzymes. These proteins are both essential for the replication		
CC	of the virus, acting to cleave its replicative proteins from the		
CC	polyprotein produced from the HCV genome. Inhibitors of the two proteins		
CC	should be effective as antiviral treatments of HCV infection. This is		
CC	useful as HCV can lead to chronic liver disease such as cirrhosis, liver		
CC	failure and liver cancer. The present invention concerns a number of NS3		
CC	mutants and NS3-NS4A fusion proteins which can be used to identify		
CC	inhibitors of this type, as well as enabling structural studies of the		
CC	protease and protease:inhibitor complexes. The protein produced from this		
CC	sequence contains the alpha-helix0-1 variant.		
XX			
SQ	Sequence 588 BP; 103 A; 180 C; 156 G; 149 T; 0 other;		

Alignment Scores:	
Pred. No.:	5,38e-78
Score:	929.00
Percent Similarity:	94.42%
Best Local Similarity:	92.89%
Query Match:	91.44%
DB:	21
	2
	1
	1
Length:	588
Matches:	183
Conservative:	3
Mismatches:	9
Indels:	2
Gaps:	1

NS-09-965-594-22 (1-197) x AAA73329 (1-588)

Qy	1	MetLysLysGlySerValIleValGlyArgIleAsnLeuSerGlyAspThrAla	20
Db	1	ATCAAAAAAAAAAGGATCGGTGTTATCTCGCGCGGTATAGTACTGAACGGT-----GCT	54
Qy	21	TyrAlaGlnGlnThrArgGlyGlnGlnGlyThrGlnLysThrSerHisThrGlyArgAsp	40
Db	55	TAGCTCACACACCTCGAGGTGAGAGGGTTCCTCAAGNAACCTCCACACCGGTCTGTGAC	114
Qy	41	LysAsnGlnValGlnGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla	60
Db	115	AAAAACACAGTTGAAGGTGAAGTTCAGATCGTTTCCACCGCTGCTCAGACCTTCTCGGCT	174
Qy	61	ThrSerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla	80
Db	175	ACCTGCATCAACGGTGTGCTGAGACGGTTACCAACGGTCTGCTACCGGTACCATCGCT	234
Qy	81	SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr	100
Db	235	TCGCCAAAAGGTCGGGTTATCCAGATGTATCCAGTGTGACACCAACGTTGACAAAGACCTGTGGTTGG	294
Qy	101	GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr	120
Db	295	CGGGCTCCGACGGTTCGGTTCCCTGACCCCGTGCACCTCGGGTTCCTCCGACCTGTGAC	354
Qy	121	LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer	140
Db	355	CTGGTTACCCGTCACGCTGACGTTATCCGGTTTCGCTCGTGGTGACTCCCGTGGGTTC	414
Qy	141	LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys	160
Db	415	CTGCTGCCCGCGTCCGATCTCTACCTGAAGGTTCCTCCGGTGGTCCGCGTCTGTGC	474

QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180
 DB 475 CCGGTGGTCACGCTGTTGGTATCTTCGTCGTGTTGCACCCGGGGTGGCTAAA 534
 QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
 DB 535 GCTGTTGACTTCATCCCGGTTGAATCCCTGGAACACCATGCGTTCCCGG 585

RESULT 8
 AAA73328
 ID AAA73328 standard; DNA; 588 BP.
 AC AAA73328;
 XX
 XX
 DT 19-DEC-2000 (first entry)
 XX
 DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #1.
 XX
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
 KW liver failure; liver cancer; ds.
 XX
 OS Hepatitis C virus.
 OS Synthetic.
 XX
 XX
 FH Key Location/Qualifiers
 FT CDS 1..588
 FT /tag= a
 FT /product= "NS3-NS4A fusion protein"
 XX
 XX W0200040707-AL.
 XX
 XX 13-JUL-2000.
 XX
 XX 06-JAN-2000; 2000WO-US00345.
 XX
 XX 08-JAN-1999; 99US-0115271.
 XX
 XX (BRIM) BRISTOL-MYERS SQUIBB CO.
 PA
 PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
 XX
 XX WPI: 2000-465976/40.
 DR P-PSDB; AAB15212.
 XX
 XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 PT
 XX
 XX Disclosure; Fig 10; 66pp; English.
 XX
 CC The present sequence is the coding sequence for a fusion protein created
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
 CC proteins are both essential for the replication of the virus, acting to
 CC cleave its replicative proteins from the polyprotein produced from the
 CC HCV genome. Inhibitors of the two proteins should be effective as
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A
 CC fusion proteins which can be used to identify inhibitors of this type, as
 CC well as enabling structural studies of the protease and
 CC protease:inhibitor complexes.
 XX
 SQ Sequence 588 BP; 97 A; 183 C; 153 G; 155 T; 0 other;

Alignment Scores:
 Pred. No.: 1,81e-75 Length: 588
 Score: 902.00 Matches: 181
 Percent Similarity: 92.39% Conservative: 1
 Best Local Similarity: 91.88% Mismatches: 13
 Query Match: 88.78% Indels: 2
 DB: 21 Gaps: 1

US-09-965-594-22 (1-197) x AAA73328 (1-588)
 QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
 DB 1 ATGAAAAAAGGTTCCCGTTGTTATCGTCGGCGGTATAGTACTGAACGGT-----GCT 54
 QY 21 TyrAlaGlnGlnThrArgGlyGluGlnGlnThrGlnLysThrSerHisThrGlyArgAsp 40
 DB 55 TAGCTCAGCAGACTCGAGGCTGCTGGGTGCTGATCATCACCCTCCCTGACCGGTGCTGAC 114
 QY 41 LysAsnGlnValGluGluGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
 DB 115 AAAAACCCAGGTTGAAGGTGAAGTTCAGATCGTTTCCACCGCTGCTCAGACCTTCCTGGCT 174
 QY 61 ThrSerIleAsnGlyValLeuThrValThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
 DB 175 ACCTGCATCAACGGTGTTCCTGACCGGTTTACACCGGTGCTGCTACCCGTACCATCGCT 234
 QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr 100
 DB 235 TCCCGAAAGGTCGGTTATCCAGATGTACACCAACGTTGACAAAGACCTGGTGTGGTGG 294
 QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
 DB 295 CCGGCTCCGCGAGGTTCCCGTTCCTGACCCCGGTGACCTGCGGTTCCTCCGACCTGTAC 354
 QY 121 LeuValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySer 140
 DB 355 CTGGTTACCGCTCACGCTGACGTTATCCCGGTTCCGTCGTCGTGCTGACCTCCCGGTGCC 414
 QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
 DB 415 CTGCTGTCCCGCGTCGATCTCCTACTGAAAGGTTCCCTCCGCTGCTCCGCTGCTGTCG 474
 QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180
 DB 475 CCGGCTGCTCAGCTGTTGGTATCTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTA 534
 QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
 DB 535 GCTGTTGACTTCATCCCGGTTGAATCCCTGGAACACCATGCGTTCCCGG 585

RESULT 9
 ABA95615
 ID ABA95615 standard; DNA; 12734 BP.
 XX
 AC ABA95615;
 XX
 DT 21-MAR-2002 (first entry)
 XX
 DE Chimeric BVDV/HCV NS3-wt sequence.
 XX
 KW Pestivirus; Npro; protease; NS3; screening; ds.
 XX
 OS Chimeric - Bovine viral diarrhea virus.
 OS Chimeric - Hepatitis C virus.
 XX
 PN USG326137-B1.
 XX
 PD 04-DEC-2001.
 XX
 XX 25-JUN-1999; 99US-0344456.
 XX
 PR 25-JUN-1999; 99US-0344456.
 XX
 PA (SCHE) SCHERING CORP.
 XX
 PI Hong 2, Lai VCH, Lau JYN;
 XX
 DR WPI; 2002-121103/16.
 XX
 PT Nucleic acid construct encoding chimeric Hepatitis C Virus (HCV)

PT pestivirus genome where the Npro protease gene is replaced with NS3
 PT protease gene, useful for in vivo screening of compounds which inhibit
 PT HCV infection

XX Example 2: Columns 17-28; 20pp; English.

XX The present invention relates to a nucleic acid construct encoding a
 CC chimeric Hepatitis C virus (HCV)-pestivirus genome. The construct
 CC comprises a pestivirus genome where a Npro pestivirus protease gene is
 CC replaced with a gene encoding a functional HCV NS3 protease. Furthermore,
 CC each junction site recognised by the Npro protease is replaced with a
 CC junction site recognised by the HCV NS3 protease. The construct is useful
 CC for screening compounds that inhibit HCV in vivo by inhibiting HCV
 CC protease, where screening may be in cell culture or in an animal model.
 CC The present sequence is a chimeric clone of BVDV (bovine viral diarrhoea
 CC virus)/HCV NS3-wt, which was used to illustrate the present invention.

XX Sequence 12734 BP; 4032 A; 2604 C; 3295 G; 2803 T; 0 other;

Alignment Scores:
 Pred. No.: 5,54e-72 Length: 12734
 Score: 882.50 Matches: 176
 Percent Similarity: 92.31% Conservative: 4
 Best Local Similarity: 90.26% Mismatches: 12
 Query Match: 86.86% Indels: 3
 DB: 24 Gaps: 1

US-09-965-594-22 (1-197) x ABA95615 (1-12734)

QY 5 GlySerValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
 DB 413 GGTAGTGTGTATTGTTGGTAGAATGTTTNTCTGGTAGTGTGTATCATCAGCGGTAC 472
 QY 22 AlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAspLys 41
 DB 473 GCCCAGCAGCAGAGAGGCTCTAGGGTGAAGATCACCACTGACTGCTGGCGGGACAAA 532
 QY 42 AsnGlnValGluGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
 DB 533 AACCAAGTGGAGGGTGGAGTCCAGATGCTCACTGCCAACCTTCTCGCAACG 592
 QY 62 SerIleAsnGlyValLeuTyrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
 DB 593 TGCATCAATGGGTATGCTGGAGTCTTACCAAGCGGGCCGACGAGCACCATCGATCA 652
 QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrPhe 101
 DB 653 CCCAAGGGTCTCTGTCATCCAGATGTATACCAATGTGGACCAAGACCTTGTGGGTGCC 712
 QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
 DB 713 GCTCTCTCAAGTTCCTCCCTCATGTACACCTCGACCTCGGCTCTCTCGSACCTTTACCTG 772
 QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyArgGlyAspSerArgGlySerLeu 141
 DB 773 GTTACGAGCAGCGCAGCTATCTCCGTCGCCGCGGAGGTATAGCAGGGGTAGCCTG 832
 QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
 DB 833 CTTTCGCCCGGCCATTTCTTACCTAAAGGCTCTCGGGGGGTCCCGCTGTGTGCCCC 892
 QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
 DB 893 GCGGGACACCGCTGGCCCTATTACGGCCCGGGGTGTGCACCCGTGGAGTGCCCAAGCG 952
 QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196
 DB 953 GTGGACTTTATCCCTGTGGAGAACCTAGAGCAACCATGAGATCC 997

RESULT 10

AA80355

ID AAX80355 standard; cDNA; 1998 BP.

XX

AC AAX80355;
 XX 07-SEP-1999 (first entry)
 DT
 DE HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:105.

XX HCV; hepatitis C virus; single chain recombinant complex; linker;
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
 KW hydrophobic domain; covalent complex; detection; inhibitor; ss.

OS Hepatitis C virus.

OS Synthetic.

XX WO9928482-A2.

XX 10-JUN-1999.

XX 24-NOV-1998; 98WO-US24528.

XX 28-JUL-1998; 98US-0094331.

XX 28-NOV-1997; 97US-0067315.

XX (SCHE) SCHERING CORP.

XX Malcolm BA, Taremi SS, Weber PC, Yao N;

XX WPI; 1999-385385/32.

XX New hepatitis C virus covalent complexes

XX Disclosure; Page 166-169; 21lpp; English.

XX The present invention describes a covalent hepatitis C virus (HCV)
 CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
 CC to the amino terminus of the HCV NS3 protease domain. The present
 CC sequence encodes an example of the above complex. The covalent
 CC NS4A-NS3 complexes are useful for structural determination and
 CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.
 CC They can also be used for detecting inhibitors of the protease activity,
 CC the helicase activity and the ATPase activity of NS3. The covalent
 CC NS4A-NS3 complexes are more soluble, stable and active than the non-
 CC covalent protease-peptide complexes previously available.

XX Sequence 1998 BP; 411 A; 595 C; 569 G; 423 T; 0 other;

Alignment Scores:
 Pred. No.: 2.5e-72 Length: 1998
 Score: 875.50 Matches: 167
 Percent Similarity: 92.86% Conservative: 15
 Best Local Similarity: 85.20% Mismatches: 11
 Query Match: 86.17% Indels: 3
 DB: 20 Gaps: 1

US-09-965-594-22 (1-197) x AAX80355 (1-1998)

QY 5 GlySerValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
 DB 64 GGTCTGTGTTGTTATTGTTGGTAGAATATTATCTGGTAGTGGTAGTATCAGCGCTAC 123
 QY 22 AlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAspLys 41
 DB 124 TCCCAACAGACGCGGGGCTACTTGTTCGAAGAAGACTAGCTTACAGCGCGGACAA 183
 QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
 DB 184 AACCAAGTGGAGGGTTCAGGTTCACCGCGCAACAAATCCTTCTCGCGACC 243

RESULT 10

AA80355

ID AAX80355 standard; cDNA; 1998 BP.

XX

QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101
DB 304 CCAAGGGGCCAATCACCAGATGTACACTAATGTGGACAGGACCTGCTGGCTGGCAG 363
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
DB 364 GCGCCCGCGGGCGGTCTCTGACACCATGCCCTGTGGCAGCTCAGACCTTTACTTG 423
QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
DB 424 GTACGAGACATGTGAGCTCATTCGGTGGCGCGGGGGGAGAGCTGAGGGGAGCTG 483
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerGlyProLeuLeuCysPro 161
DB 484 CTCCTCCCGCAGCCCTGCTCTCTACTTTGAAGGCTCTTCGGGTGGTCCACTGCTGCCCT 543
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
DB 544 TCGGGGACAGCTGTGGGCATCTTCGGGCTGCCGTATGCACCCGGGGGTTCGGAAGCG 603
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
DB 604 GTGGACTTGTGCGCGTAGAGTCCATGGGAACACTACTATCGCGTCTCCG 651
RESULT 11
ID AAX80359 standard; cDNA: 1998 BP.
XX AAX80359;
XX AAX80359;
XX 07-SEP-1999 (first entry)
XX HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:109.
XX HCV; hepatitis C virus; single chain recombinant complex; linker;
XX NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
XX hydrophobic domain; covalent complex; detection; inhibitor; ss.
XX Hepatitis C virus.
XX Synthetic.
XX WO9928482-A2.
XX 10-JUN-1999.
XX 24-NOV-1998; 98WO-US24528.
XX 28-JUL-1998; 98US-0094331.
XX 28-NOV-1997; 97US-0067315.
XX (SCHE) SCHERING CORP.
XX Malcolm BA, Taremi SS, Weber PC, Yao N;
XX WPI; 1999-385385/32.
XX New hepatitis C virus covalent complexes
XX Disclosure; Page 179-182; 21lpp; English.
XX The present invention describes a covalent hepatitis C virus (HCV)
XX NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
XX NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
XX hydrophobic domain of native HCV NS4A peptide is tethered by the linker
XX to the amino terminus of the HCV NS3 protease domain. The present
XX sequence encodes an example of the above complex. The covalent
XX NS4A-NS3 complexes are useful for structural determination and
XX determination of mode of binding of HCV inhibitors by NMR spectroscopy.
XX They can also be used for detecting inhibitors of the protease activity,
XX the helicase activity and the ATPase activity of NS3. The covalent
XX NS4A-NS3 complexes are more soluble, stable and active than the non-
XX covalent protease-peptide complexes previously available.

SQ Sequence 1998 BP; 411 A; 595 C; 569 G; 423 T; 0 other;
Alignment Scores:
Pred. No.: 4,77e-72 Length: 1998
Score: 872.50 Matches: 166
Percent Similarity: 92.86% Conservative: 16
Best Local Similarity: 84.69% Mismatches: 11
Query Match: 85.88% Indels: 3
DB: 20 Gaps: 1
US-09-965-594-22 (1-197) x AAX80359 (1-1998)
QY 5 GlySerValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
DB 64 GGTTCGTGTTTATCTGTGTAGAAATATTTATCTGGTAGTATCATCGGCTTAC 123
QY 22 AlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAspLys 41
DB 124 TCCCAACAGACGCGGGGCGCTACTTGGTTGCAAGAGACTAGCCTTACAGCGCGGACAG 183
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
DB 184 AACGAGTCGAGGAGAGGTTTCAGGTGTTCCACCGCAACACAAATCCTTCTCGGCGAC 243
QY 62 SerIleAsnGlyValLeuThrPheValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
DB 244 TGGGTCAACGGCGTGTGTGGACCGTTTACCATGTGTGGCTCAAGACCTTAGCGCGC 303
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101
DB 304 CCAAGGGGCCAATCACCAGATGTACACTAATGTGGACAGGACCTGCTGGCTGGCAG 363
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
DB 364 GCGCCCGCGGGCGGTCTCTTGCACACCATGCCCTGTGGCAGCTCAGACCTTTACTTG 423
QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
DB 424 GTACGAGACATGCTGACGTCATTCGGTGGCGCGGGGGGACAGTAGGGGAGCGCTG 483
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
DB 484 CTCCTCCCGCAGCCCTGCTCTCTACTTTGAAGGCTCTTCGGGTGGTCCACTGCTGCCCT 543
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
DB 544 TCGGGGACAGCTGTGGGCATCTTCGGGCTGCCGTATGCACCCGGGGGTTCGGAAGCG 603
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
DB 604 GTGGACTTGTGCGCGTAGAGTCCATGGGAACACTACTATCGCGTCTCCG 651
RESULT 12
AAX80354
ID AAX80354 standard; cDNA: 1998 BP.
XX AAX80354;
XX AAX80354;
XX 07-SEP-1999 (first entry)
XX HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:104.
XX HCV; hepatitis C virus; single chain recombinant complex; linker;
XX NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
XX hydrophobic domain; covalent complex; detection; inhibitor; ss.
XX Hepatitis C virus.
XX Synthetic.
XX WO9928482-A2.
XX 10-JUN-1999.

PF 24-NOV-1998; 98WO-US24528.
XX 28-JUL-1998; 98US-0094331.
PR 28-NOV-1997; 97US-0067315.
XX (SCHE) SCHERING CORP.
XX Malcolml BA, Taremi SS, Weber PC, Yao N;
PI WPI; 1999-385385/32.
DR New hepatitis C virus covalent complexes
XX Disclosure; Page 163-166; 211pp; English.
XX The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence encodes an example of the above complex. The covalent
CC NS4A-NS3 complexes are useful for structural determination and
CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.
CC They can also be used for detecting inhibitors of the protease activity,
CC the helicase activity and the ATPase activity of NS3. The covalent
CC NS4A-NS3 complexes are more soluble, stable and active than the non-
CC covalent protease-peptide complexes previously available.
XX Sequence 1998 BP; 410 A; 596 C; 568 G; 424 T; 0 other;
SQ Alignment Scores:
Pred. No.: 5,91e-72 Length: 1998
Score: 871.50 Matches: 167
Percent Similarity: 92.35% Conservative: 14
Best Local Similarity: 85.20% Mismatches: 12
Query Match: 85.78% Indels: 3
DB: 20 Gaps: 1
US-09-965-594-22 (1-197) x AAX80354 (1-1998)
QY 5 GlySerValIleValGluValGlnLeuSerGlyAsp-----ThrAlaTyr 21
DB 64 GTTCTGTTGTTATTTGTTAGATATTTTCTGTTAGTGTAGTATCATCGGCTAC 123
QY 22 AlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAspLys 41
DB 124 TCCCAACAGACGCGGGGCTACTTGGTTCATCAAGACTACGCTTACAGCGCGGACAAG 183
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
DB 184 AACCAAGTTCGAGGAGAGGTTTCAGTGTGTTCCACCGCAACACATCTCTCTGCGGACC 243
QY 62 SerIleAsnGlyValLeuThrPheValThrHisGlyAlaGlyThrArgThrIleAlaSer 81
DB 244 TCGCTCAACGCGGCTGTGTGGACCGTTTACCATGCTGCTCTCAAGACCTTAGCCGCG 303
QY 82 ProLysGlyProValThrGlnMetThrThrAsnValAspLysAspLeuValGlyTyrGln 101
DB 304 CCAAGAGGCGCAATCACCATGATGATACATATGTTGACGAGGACCTCGTGGCTGGCAG 363
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
DB 364 GCGCCCCCGGGCGGCTCTCTGCACCATGTCACCTGTGCGCTAGACCTTTACTTG 423
QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
DB 424 GTCAGGAGACATGCTGACGTCATTCGCTGCGCGCGGCGGCGGACAGTAGGGGAGCGCTG 483
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
DB 484 CFTCTCCCGAGCGCTCTCTCTACTTGAAGGGCTCTTCGGGTGTCTCCACTCTCTGCGCT 543
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181

DB 544 TCGGGGCACGCTGTGGCATCTTCGGGCTGCGGTATGCACCCGGGGTTCGGAAGCG 603
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
DB 604 GTGGACTTTGTGCCGTAGATCCATGGAACACTACTATGCGGTCTCCG 651
RESULT 13
AAX80345
ID AAX80345 standard; cDNA; 651 BP.
XX AAX80345;
XX 07-SEP-1999 (first entry)
XX HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:95.
XX HCV; hepatitis C virus; single chain recombinant complex; linker;
KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
KW hydrophobic domain; covalent complex; detection; inhibitor; ss.
XX Hepatitis C virus.
OS Synthetic.
XX WO9928482-A2.
XX 10-JUN-1999.
XX 24-NOV-1998; 98WO-US24528.
XX 28-JUL-1998; 98US-0094331.
PR 28-NOV-1997; 97US-0067315.
XX (SCHE) SCHERING CORP.
XX Malcolml BA, Taremi SS, Weber PC, Yao N;
PI WPI; 1999-385385/32.
XX New hepatitis C virus covalent complexes
PT Disclosure; Page 147-148; 211pp; English.
XX The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence encodes an example of the above complex. The covalent
CC NS4A-NS3 complexes are useful for structural determination and
CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.
CC They can also be used for detecting inhibitors of the protease activity,
CC the helicase activity and the ATPase activity of NS3. The covalent
CC NS4A-NS3 complexes are more soluble, stable and active than the non-
CC covalent protease-peptide complexes previously available.
XX Sequence 651 BP; 120 A; 187 C; 200 G; 144 T; 0 other;
SQ Alignment Scores:
Pred. No.: 2.8e-72 Length: 651
Score: 868.50 Matches: 166
Percent Similarity: 92.82% Conservative: 15
Best Local Similarity: 85.13% Mismatches: 11
Query Match: 85.48% Indels: 3
DB: 20 Gaps: 1
US-09-965-594-22 (1-197) x AAX80345 (1-651)
QY 5 GlySerValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
DB 64 GTTCTGTTGTTATTTGTTAGATATTTTCTGTTAGTGTAGTATCATCGGCTAC 123
QY 22 AlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAspLys 41

```

Db 124 TCCCAACAGACGCGGGGCTACTTGGTTCCAAAGAGACTACGCTTACAGCGCGGCACAAAG 183
Qy 42 AsnGlnValGluGlyValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
Db 184 AACCAAGTCGAGGAGAGGTTGAGTGGTTCCACCCGCAACACAAATCTCTCTGGCGACC 243
Qy 62 SerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
Db 244 TCGCGTCAACGGCGTGTGTGGACCGTTTACCATGTGTGCTGGCTCAACAGACCTTAGCGCGC 303
Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
Db 304 CCAAGAGGGCCCAATCACCCAGATGTACACTAATGTGGACCAAGGACCTCGTGGCTGGCAG 363
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
Db 364 GGGCCCCCGGGCGGCTTCTTGACACCATGCCTGTGGCAGCTCAGACCTTACTTG 423
Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
Db 424 GTCAGGACATGCTGACGCTATTCGGGTGCGCGCGGCGGCGACAGTAGGGGAGCCTG 483
Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
Db 484 CTCTCCCGCCAGCGCTGCTCCTACTTGAAGGGCTCTGCTGGTGTGCCACTCTCTGCCCT 543
Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
Db 544 TCGGGGCACGCTGTGGGCATCTTCCGGCTGCGCTATGCACCCGGGGGGTTCGAGGCG 603
Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
Db 604 GTGGACTTTGTGGCGGTAGATCCATGGAAACTACTATGCGGCTCTCCG 651

RESULT 15
AAx80358
ID AAX80358 standard; cDNA; 1998 BP.
XX
AC AAX80358;
XX
DT 07-SEP-1999 (first entry)
XX
DE HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:108.
XX
KW HCV; hepatitis C virus; single chain recombinant complex; linker;
KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
KW hydrophobic domain; covalent complex; detection; inhibitor; ss.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
PN W09928482-A2.
XX
PD 10-JUN-1999.
XX
PF 24-NOV-1998; 98NO-US24528.
XX
PR 28-JUL-1998; 98US-0094331.
XX
PR 28-NOV-1997; 97US-0067315.
XX
PA (SCHE ) SCHERING CORP.
XX
PI Malcolm BA, Taremi SS, Weber PC, Yao N;
XX
WIPI; 1999-385385/32.
XX
PT New hepatitis C virus covalent complexes
XX
PS Disclosure; Page 176-179; 21pp; English.
XX
CC The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV

```

```

CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence encodes an example of the above complex. The covalent
CC NS4A-NS3 complexes are useful for structural determination and
CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.
CC They can also be used for detecting inhibitors of the protease activity,
CC the helicase activity and the ATPase activity of NS3. The covalent
CC NS4A-NS3 complexes are more soluble, stable and active than the non-
CC covalent protease-peptide complexes previously available.
XX
SQ Sequence 1998 BP; 410 A; 596 C; 568 G; 424 T; 0 other;
Alignment Scores:
Pred. No.: 1,13e-71 Length: 1998
Score: 868.50 Matches: 166
Percent Similarity: 92.35% Conservative: 15
Best Local Similarity: 84.69% Mismatches: 12
Query Match: 85.48% Indels: 3
DB: 20 Gaps: 1
US-09-965-594-22 (1-197) x AAX80358 (1-1998)
Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
Db 64 GGTTCCTGTTGTTATGTTGTTAGTAATTTATTTATCTGTTAGTGTAGTATCATCGGCTAC 123
Qy 22 AlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAspLys 41
Db 124 TCCCAACAGACGCGGGGCTACTTGGTTGTCATCAAGACTAGCCTTACAGCGCGGCACAA 193
Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
Db 184 AACCAAGTCGAGGAGAGGTTTCCAGTGGTTTCCACCGCAACACAAATCTCTCTGGCGACC 243
Qy 62 SerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
Db 244 TCGCGTCAACGGCGTGTGTGGACCGTTTACCATGTGTGCTGCTCAAGACCTTAGCGCGC 303
Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
Db 304 CCAAGAGGGCCCAATCACCCAGATGTACACTAATGTGGACCAAGGACCTCGTGGCTGGCAG 363
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
Db 364 GGGCCCCCGGGCGGCTTCTTGACACCATGCCTGTGGCAGCTCAGACCTTACTTG 423
Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
Db 424 GTCAGGACATGCTGACGCTATTCGGGTGCGCGCGGCGGCGACAGTAGGGGAGCCTG 483
Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
Db 484 CTCTCCCGCCAGCGCTGCTCCTACTTGAAGGGCTCTGCTGGTGTGCCACTCTCTGCCCT 543
Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
Db 544 TCGGGGCACGCTGTGGGCATCTTCCGGCTGCGCTATGCACCCGGGGGGTTCGAGGCG 603
Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
Db 604 GTGGACTTTGTGGCGGTAGATCCATGGAAACTACTATGCGGCTCTCCG 651

RESULT 15
AAx80353
ID AAX80353 standard; cDNA; 1998 BP.
XX
AC AAX80353;
XX
DT 07-SEP-1999 (first entry)
XX
DE HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:103.
XX

```

KW HCV; hepatitis C virus; single chain recombinant complex; linker;
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
 KW hydrophobic domain; covalent complex; detection; inhibitor; ss.

OS Hepatitis C virus.
 OS Synthetic.

XX W09928482-A2.

PN 10-JUN-1999.

XX 24-NOV-1998; 98WO-US24528.

XX 28-JUL-1998; 98US-0094331.

PR 28-NOV-1997; 97US-0067315.

XX (SCHE) SCHERING CORP.

PA Malcolm RA, Taremi SS, Weber PC, Yao N;

XX WPI; 1999-385385/32.

XX New hepatitis C virus covalent complexes

PS Disclosure; Page 160-162; 21pp; English.

XX The present invention describes a covalent hepatitis C virus (HCV)

CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV

CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the

CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker

CC to the amino terminus of the HCV NS3 protease domain. The present

CC sequence encodes an example of the above complex. The covalent

CC NS4A-NS3 complexes are useful for structural determination and

CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.

CC They can also be used for detecting inhibitors of the protease activity,

CC the helicase activity and the ATPase activity of NS3. The covalent

CC NS4A-NS3 complexes are more soluble, stable and active than the non-

CC covalent protease-peptide complexes previously available.

XX Sequence 1998 BP; 410 A; 596 C; 568 G; 424 T; 0 other;

XX

SQ

Alignment Scores:

Pred. No.: 1 4e-71 Length: 1998

Score: 867.50 Matches: 166

Percent Similarity: 92.35% Conservative: 15

Best Local Similarity: 84.69% Mismatches: 12

Query Match: 85.38% Indels: 3

DB: 20 Gaps: 1

US-09-965-594-22 (1-197) x AAX80353 (1-1998)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21

DB 64 GGTCTGTGTGTTATTTGGTAGAATATTTATCTGGTAGTGTATACCGGCTAC 123

QY 22 AlaGlnGlnThrArgGlyGluGlnGlyThrGlnIleValSerHisThrGlyArgAspLys 41

DB 124 TCCACACACACCGGGGCTTCTGTTGCAAGATCCTAGCTTACAGCGGGGACAAG 183

QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61

DB 184 AACAGGTCGAGGAGAGGTTTCAGGTGTTCCACCGCAACACAACTCTCTCGCGGACC 243

QY 62 SerIleAsnGlyValLeuThrValThrValThrHisGlyAlaGlyThrArgThrIleAlaSer 81

DB 244 TCGTCAACCGGCTGTGTTGGACCCCTTTACCATGCTGCTCAAGACCTTAGCCGGC 303

QY 82 ProLysGlyProValThrGlnMetThrAsnValAspLysAspLeuValGlyTyrPheGln 101

DB 304 CCNAAAGGGCCCAATACCCAGATGTACACTAATGTGGACAGACCTCTCGGCTGCAG 363

QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121

DB

Db 364 CGGCCCCCGGGGGCGCTTCTTGCACCATGCACCTGTGCAGCTCAGACCTTTACTTG 423

Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141

Db 424 GTCACGAGACATGCTGACGTCATTCGGTGCAGCGGGGGCCACAGTAGGGGGAGCCTG 483

Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161

Db 484 CTCCTCCCCCAGGCTGTCTCTACTTGAAGGGCTCTTCGGGTGGTCCACTGCTCTCCCT 543

Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181

Db 544 TCGGGGCACGCTGGGCATCTTCCGGCTGCCGTATGCACCGGGGGGTTCGAAAGCG 603

Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197

Db 604 GTGGACTTTGTGCGCGTAGAGTCCATGGAACCTACTATGCGGTCTCCG 651

Search completed: August 30, 2003, 19:48:21

Job time : 188.939 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 30, 2003, 19:26:03 ; Search time 176.482 Seconds
(without alignments)
2560.981 Million cell updates/sec

Title: US-09-965-594-22

Perfect score: 1016

Sequence: 1 MKKGSVVIVGRINLSGDTA.....VAKAVDFIPVESLTTMRSP 197

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1533700 seqs, 1147125425 residues

Total number of hits satisfying chosen parameters: 3067400

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09965594/runat_29082003_151920_28367/app_query.fasta_1.2872
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR SCORE=pct -THR MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORW=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09965594.ecgn_1.864.@runat_29082003_151920_28367
-NCPUP=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_NA.*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/Pct_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1016	100.0	594	10	US-09-965-594-23 Sequence 23, Appl

2	1010	99.4	594	10	US-09-965-594-21	Sequence 21, Appl
3	995	97.9	594	10	US-09-965-594-19	Sequence 19, Appl
4	980	96.5	594	10	US-09-965-594-17	Sequence 17, Appl
5	963	94.8	594	10	US-09-965-594-15	Sequence 15, Appl
6	936	92.1	594	10	US-09-965-594-25	Sequence 25, Appl
7	929	91.4	588	10	US-09-965-594-13	Sequence 13, Appl
8	902	88.8	588	10	US-09-965-594-4	Sequence 4, Appl
9	864.5	85.1	612	14	US-10-133-133A-6	Sequence 6, Appl
10	848.5	83.5	9646	9	US-09-742-659-3	Sequence 3, Appl
11	848.5	83.5	9646	10	US-09-238-076-1	Sequence 1, Appl
12	848.5	83.5	9646	11	US-09-995-937-1	Sequence 1, Appl
13	848.5	83.5	9646	11	US-09-917-563-1	Sequence 1, Appl
14	848.5	83.5	12980	10	US-09-238-076-5	Sequence 5, Appl
15	848.5	83.5	12980	11	US-09-995-937-5	Sequence 5, Appl
16	848.5	83.5	12980	11	US-09-917-563-5	Sequence 5, Appl
17	844.5	83.1	9379	9	US-09-916-359-1	Sequence 1, Appl
18	844.5	83.1	9416	10	US-09-238-076-19	Sequence 19, Appl
19	844.5	83.1	9416	11	US-09-995-937-19	Sequence 19, Appl
20	844.5	83.1	9416	11	US-09-917-563-19	Sequence 19, Appl
21	842	82.9	549	10	US-09-965-594-2	Sequence 2, Appl
22	842	82.9	2058	10	US-09-881-654-1	Sequence 1, Appl
23	842	82.9	2058	10	US-09-881-239-2	Sequence 2, Appl
24	841.5	82.8	836	10	US-09-921-397-120	Sequence 120, Appl
25	841.5	82.8	10803	10	US-09-747-419-17	Sequence 17, Appl
26	841.5	82.8	10803	14	US-10-259-275-17	Sequence 17, Appl
27	838.5	82.5	9416	10	US-09-929-955-13	Sequence 13, Appl
28	838.5	82.5	9416	13	US-10-104-966-13	Sequence 13, Appl
29	838	82.5	2061	10	US-09-929-955-16	Sequence 16, Appl
30	826.5	81.3	13910	11	US-09-919-901-1	Sequence 1, Appl
31	823.5	81.1	13910	11	US-09-919-901-8	Sequence 8, Appl
32	823.5	81.1	13910	11	US-09-919-901-15	Sequence 15, Appl
33	823	81.0	2064	11	US-09-884-456-69	Sequence 69, Appl
34	823	81.0	2523	11	US-09-884-456-85	Sequence 85, Appl
35	820.5	80.8	2073	14	US-10-133-133A-5	Sequence 5, Appl
36	820	80.7	6189	14	US-10-259-275-41	Sequence 41, Appl
37	820	80.7	7992	13	US-10-005-469-1	Sequence 1, Appl
38	820	80.7	7992	13	US-10-005-469-2	Sequence 2, Appl
39	820	80.7	7992	13	US-10-005-469-4	Sequence 4, Appl
40	820	80.7	7992	13	US-10-005-469-6	Sequence 6, Appl
41	820	80.7	8638	12	US-10-309-561-24	Sequence 24, Appl
42	820	80.7	8638	13	US-10-029-907-24	Sequence 24, Appl
43	820	80.7	8639	12	US-10-309-561-1	Sequence 1, Appl
44	820	80.7	8639	13	US-10-029-907-1	Sequence 1, Appl
45	820	80.7	8642	12	US-10-309-561-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1

US-09-965-594-23
; Sequence 23, Application US/09965594
; Patent No. US20020106642A1
; GENERAL INFORMATION:
; APPLICANT: Wittekind, Michael
; APPLICANT: Weinheimer, Steven
; APPLICANT: Zhang, Yaquin
; APPLICANT: Goldfarb, Valentina
; TITLE OF INVENTION: Modified Forms of Hepatitis C NS3 Protease for
; TITLE OF INVENTION: Facilitating Inhibitor Screening and Structural Studies
; FILE REFERENCE: DB17Sequences
; CURRENT APPLICATION NUMBER: US/09/965,594
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/115,271
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 23
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Hepatitis C virus
US-09-965-594-23

```
Alignment Scores:
Pred. No.: 4,57e-109 Length: 594
Score: 1016.00 Matches: 197
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-09-965-594-22 (1-197) x US-09-965-594-23 (1-594)
QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
DB 1 ATGAAAAAAGGATCGCGTGTATATCGTCGCGCGTATCAACCTGTCGGTGACACCGCT 60
QY 21 TyrAlaGlnGlnThrArgGlyCluGlnGlnGlyThrSerHisThrGlyArgAsp 40
DB 61 TACGCTCAGCAGCTCGAGGTGAGGAGGTACCCAGAGAGCTCCACACCGGTGTCGAC 120
QY 41 LysAsnGlnValGluGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
DB 121 AAAAAACAGGTGAAGTTCAGATCGTTTCCACCGCTACCCAGACCTTCCCTGGCT 180
QY 61 ThrSerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
DB 181 ACCTCCATCAACGGGTGTTCTGTGGACCGTTTACCACGGTGTGTACCGTACCATCGCT 240
QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp 100
DB 241 TCCCGAAAGGTCGGTTACCCAGATGTACACCAAGGTGACAAAGACCTGCTGGTTGG 300
QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
DB 301 CAGGCTCCAGCGGTCCCGTTCCTGACCCCGTGCACCTCGGTTCCTCCGACCTGTAC 360
QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
DB 361 CTGGTTACCGTCACGTCAGCTTATCCCGGTTCGTCGTCGTCGTCGTCGTCGTCGTC 420
QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerGlySerGlyProLeuLeuCys 160
DB 421 CTGCTGTCCCGTCGTCGTCATCTCCACCTGAAAGGTTCCTCCGGTGGTCCGCTGTCG 480
QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180
DB 481 CCGGCTGGTCAGCTGTGGTATCTCCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 540
QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
DB 541 GCTGTGTGACTTCATCCCGGTGAATCCCTGGAACACCATCGCTTCCCGG 591

RESULT 2
US-09-965-594-21
; Sequence 21, Application US/09965594
; Patent No. US20020106642A1
; GENERAL INFORMATION:
; APPLICANT: Wittekind, Michael
; APPLICANT: Weinheimer, Steven
; APPLICANT: Zhang, Yaqu
; APPLICANT: Goldfarb, Valentina
; TITLE OF INVENTION: Modified Forms of Hepatitis C NS3 Protease for
; TITLE OF INVENTION: Facilitating Inhibitor Screening and Structural Studies
; FILE REFERENCE: DB17Sequences
; CURRENT APPLICATION NUMBER: US/09/965,594
; PRIOR FILING DATE: 2001-09-27
; PRIOR FILING DATE: 60/115,271
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Hepatitis C virus

RESULT 3
US-09-965-594-19
; Sequence 19, Application US/09965594
; Patent No. US20020106642A1
; GENERAL INFORMATION:
; APPLICANT: Wittekind, Michael
; APPLICANT: Weinheimer, Steven
; APPLICANT: Zhang, Yaqu
; APPLICANT: Goldfarb, Valentina
; TITLE OF INVENTION: Modified Forms of Hepatitis C NS3 Protease for
; TITLE OF INVENTION: Facilitating Inhibitor Screening and Structural Studies
; FILE REFERENCE: DB17Sequences
; CURRENT APPLICATION NUMBER: US/09/965,594
; PRIOR FILING DATE: 2001-09-27
; PRIOR FILING DATE: 60/115,271
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 594
```


; TYPE: DNA
; ORGANISM: Hepatitis C virus
US-09-965-594-19

Alignment Scores:
Pred. No.: 1,27e-106 Length: 594
Score: 995.00 Matches: 193
Percent Similarity: 98.98% Conservative: 2
Best Local Similarity: 97.97% Mismatches: 2
Query Match: 97.93% Indels: 0
DB: 10 Gaps: 0

US-09-965-594-22 (1-197) x US-09-965-594-19 (1-594)

```
QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
   |||||
Db 1 ATGAAAAAAGGATCCGTTGTTATCGTCGCGCGTATCAACCTGTCGCGTGACACCGCT 60
   |||||
QY 21 TyrAlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAsp 40
   |||||
Db 61 TAGCGTCAAGCAGACTCAGGTGAGGAGGTGCGCAAGAAACCTCCAGACCGGTGCTGAC 120
   |||||
QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
   |||||
Db 121 AAAAACCGAGTTGAAGTGAAGTTCAGATCGTTTCACCGCTACCCAGACCTTCCTGGCT 180
   |||||
QY 61 ThrSerIleAsnGlyValLeuThrThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
   |||||
Db 181 ACCTCCATCAACGGTGTCTGTGGACCGTTTACCACGGTGTGGTACCGGTACCATCGCT 240
   |||||
QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr 100
   |||||
Db 241 TCCCGGAAAGGTCGGTTACCCAGATGTACACCAAGTTGACAAAGACCTGTTGGTTGG 300
   |||||
QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
   |||||
Db 301 CAGGTCGCCAGGTTCCGTTCCCTGACCCCGCTGACCTGCGGTCTCTCCACCTGTAC 360
   |||||
QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
   |||||
Db 361 CTGGTTACCCGTCACGCTACGTTATCCCGGTCGTCGTCGTCGTCGTCGTCGTCGTC 420
   |||||
QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
   |||||
Db 421 CTGCTGTCGCCGCGTCGATCTCTACCTGAAAGTTCCTCCGGTGGTCCGCTGCTGTC 480
   |||||
QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180
   |||||
Db 481 CCGGCTGGTCACGCTGTTGCTATCTTCCGTCGTCGTCGTCGTCGTCGTCGTCGTC 540
   |||||
QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
   |||||
Db 541 GCTGTTGACTTCACTCCCGTTGAATCCCTGGAAACCAACCATCGTTCCCGC 591
   |||||
```

RESULT 4

US-09-965-594-17

; Sequence 17, Application US/09965594
; Patent No. US20020106642A1
; GENERAL INFORMATION:
; APPLICANT: Wittek, Michael
; APPLICANT: Weinheimer, Steven
; APPLICANT: Zhang, Yaqu
; APPLICANT: Goldfarb, Valentina
; TITLE OF INVENTION: Modified Forms of Hepatitis C NS3 Protease for
; Facilitating Inhibitor Screening and Structural Studies
; TITLE OF INVENTION: Facilitating Inhibitor Screening and Structural Studies
; FILE REFERENCE: DB17Sequences
; CURRENT APPLICATION NUMBER: US/09/965, 594
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/115,271
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 17
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Hepatitis C virus
US-09-965-594-17

Alignment Scores:
Pred. No.: 7,06e-105 Length: 594
Score: 980.00 Matches: 190
Percent Similarity: 97.46% Conservative: 2
Best Local Similarity: 96.45% Mismatches: 5
Query Match: 96.46% Indels: 0
DB: 10 Gaps: 0

US-09-965-594-22 (1-197) x US-09-965-594-17 (1-594)

```
QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
   |||||
Db 1 ATGAAAAAAGGATCCGTTGTTATCGTCGCGCGTATCAACCTGTCGCGTGACACCGCT 60
   |||||
QY 21 TyrAlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAsp 40
   |||||
Db 61 TAGCGTCAAGCAGACTCAGGTGAGGAGGTGCGCAAGAAACCTCCAGACCGGTGCTGAC 120
   |||||
QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
   |||||
Db 121 AAAAACCGAGTTGAAGTGAAGTTCAGATCGTTTCCACCGCTACCCAGACCTTCCTGGCT 180
   |||||
QY 61 ThrSerIleAsnGlyValLeuThrThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
   |||||
Db 181 ACCTCCATCAACGGTGTCTGTGGACCGTTTACCACGGTGTGGTACCGGTACCATCGCT 240
   |||||
QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr 100
   |||||
Db 241 TCCCGGAAAGGTCGGTTACCCAGATGTACACCAAGTTGACAAAGACCTGTTGGTTGG 300
   |||||
QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
   |||||
Db 301 CAGGTCGCCAGGTTCCGTTCCCTGACCCCGCTGACCTGCGGTCTCTCCAGCTGTAC 360
   |||||
QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
   |||||
Db 361 CTGGTTACCCGTCACGCTACGTTATCCCGGTCGTCGTCGTCGTCGTCGTCGTCGTC 420
   |||||
QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
   |||||
Db 421 CTGCTGTCGCCGCGTCGATCTCTACCTGAAAGTTCCTCCGGTGGTCCGCTGCTGTC 480
   |||||
QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180
   |||||
Db 481 CCGGCTGGTCACGCTGTTGCTATCTTCCGTCGTCGTCGTCGTCGTCGTCGTCGTC 540
   |||||
QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
   |||||
Db 541 GCTGTTGACTTCACTCCCGTTGAATCCCTGGAAACCAACCATCGTTCCCGC 591
   |||||
```

RESULT 5

US-09-965-594-15

; Sequence 15, Application US/09965594
; Patent No. US20020106642A1
; GENERAL INFORMATION:
; APPLICANT: Wittek, Michael
; APPLICANT: Weinheimer, Steven
; APPLICANT: Zhang, Yaqu
; APPLICANT: Goldfarb, Valentina
; TITLE OF INVENTION: Modified Forms of Hepatitis C NS3 Protease for
; Facilitating Inhibitor Screening and Structural Studies
; TITLE OF INVENTION: Facilitating Inhibitor Screening and Structural Studies
; FILE REFERENCE: DB17Sequences
; CURRENT APPLICATION NUMBER: US/09/965, 594
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/115, 271
; SOFTWARE: PatentIn Ver. 2.0

; NUMBER OF SEQ ID NOS: 26
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 15
 ; LENGTH: 594
 ; TYPE: DNA
 ; ORGANISM: Hepatitis C virus
 US-09-965-594-15

Alignment Scores:

Pred. No.: 6,71e-103 Length: 594
 Score: 963.00 Matches: 187
 Percent Similarity: 95.94% Conservativeness: 2
 Best Local Similarity: 94.92% Mismatches: 8
 Query Match: 94.78% Indels: 0
 DB: 10 Gaps: 0

US-09-965-594-22 (1-197) x US-09-965-594-15 (1-594)

```

Qy 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
Db 1 ATGAATAAAGGATCGTGTGTTATCGTCGGCGGTATCAACCTGTCGGTGACACCGCT 60
Qy 21 TyrAlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAsp 40
Db 61 TACGCTCAGCAGACTCGAGGTGAGAGGGTTGCCAAGAAACCTCCACAGCCGGTCGTGAC 120
Qy 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
Db 121 AAAAAACAGGTGAAGTGAAGTTGAGTTCAGATCGTTTCCACCGCTGCTCAGACCTTCCGTGCT 180
Qy 61 ThrSerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
Db 181 ACCTGCATCAACGGTGTGTTGCTGGACCGTTTACCACGGTGTGTTACCCGTACCATCGCT 240
Qy 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp 100
Db 241 TCCCGAAGGTGCCGGTTATCCAGATGACCAACGGTTGACAAAGACCTGGTGGTGG 300
Qy 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
Db 301 CCGGCTCCGACGGTTCGGTTCCTGACCGCGTGCACCTCGGTTCTCCGACCTGTAC 360
Qy 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
Db 361 CTGGTTACCGTCACGCTGACGTTATCCCGGTTCGTCTGTTGACCTCCGGTGGTTC 420
Qy 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
Db 421 CTGCTGTCGCCCGCTCCGATCCTACCTGAAAGGTTCTCCGTTGGTCCGCTGCTGC 480
Qy 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180
Db 481 CCGGCTGGTCACGCTGTGGTATCTCCGCTGCTGCTGTTGCACCCGGTGGTGGCTAAA 540
Qy 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
Db 541 GCTGTTGACTTCATCCCGTTGAATCCCTGGAAACCAACCATGCGTTCCTCCCG 591

```

RESULT 6

US-09-965-594-25

; Sequence 25, Application US/09965594
 ; Patent No. US20020106642A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wittekind, Michael
 ; APPLICANT: Weinheimer, Steven
 ; APPLICANT: Zhang, Yaqu
 ; APPLICANT: Goldfarb, Valentina
 ; TITLE OF INVENTION: Modified Forms of Hepatitis C NS3 Protease for
 ; TITLE OF INVENTION: Facilitating Inhibitor Screening and Structural Studies
 ; TITLE OF INVENTION: of Protease:Inhibitor Complexes
 ; FILE REFERENCE: DB17Sequences
 ; CURRENT APPLICATION NUMBER: US/09/965,594
 ; CURRENT-FILING DATE: 2001-09-27

; PRIOR APPLICATION NUMBER: 60/115,271

; PRIOR FILING DATE: 1999-01-08

; NUMBER OF SEQ ID NOS: 26

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 25

; LENGTH: 594

; TYPE: DNA

; ORGANISM: Hepatitis C virus

US-09-965-594-25

Alignment Scores:

Pred. No.: 9.3e-100 Length: 594
 Score: 936.00 Matches: 185
 Percent Similarity: 93.91% Conservativeness: 0
 Best Local Similarity: 93.91% Mismatches: 12
 Query Match: 92.13% Indels: 0
 DB: 10 Gaps: 0

US-09-965-594-22 (1-197) x US-09-965-594-25 (1-594)

```

Qy 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
Db 1 ATGAATAAAGGATCGTGTGTTATCGTCGGCGGTATCAACCTGTCGGTGACACCGCT 60
Qy 21 TyrAlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAsp 40
Db 61 TACGCTCAGCAGACTCGAGGTGCTGGTGTGCTATCATCATCCCTCCCTGACCGGTGAC 120
Qy 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
Db 121 AAAAAACAGGTGAAGTGAAGTTGAGTTCAGATCGTTTCCACCGCTGCTCAGACCTTCCGTGCT 180
Qy 61 ThrSerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
Db 181 ACCTGCATCAACGGTGTGTTGCTGGACCGTTTACCACGGTGTGTTACCCGTACCATCGCT 240
Qy 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp 100
Db 241 TCCCGAAGGTGCCGGTTATCCAGATGACCAACGGTTGACAAAGACCTGGTGGTGG 300
Qy 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
Db 301 CCGGCTCCGACGGTTCGGTTCCTGACCGCGTGCACCTCGGTTCTCCGACCTGTAC 360
Qy 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
Db 361 CTGGTTACCGTCACGCTGACGTTATCCCGGTTCGTCTGTTGACCTCCGGTGGTTC 420
Qy 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
Db 421 CTGCTGTCGCCCGCTCCGATCCTACCTGAAAGGTTCTCCGTTGGTCCGCTGCTGC 480
Qy 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180
Db 481 CCGGCTGGTCACGCTGTGGTATCTCCGCTGCTGCTGTTGCACCCGGTGGTGGCTAAA 540
Qy 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
Db 541 GCTGTTGACTTCATCCCGTTGAATCCCTGGAAACCAACCATGCGTTCCTCCCG 591

```

RESULT 7

US-09-965-594-13

; Sequence 13, Application US/09965594

; Patent No. US20020106642A1

; GENERAL INFORMATION:

; APPLICANT: Wittekind, Michael

; APPLICANT: Weinheimer, Steven

; APPLICANT: Zhang, Yaqu

; APPLICANT: Goldfarb, Valentina

; TITLE OF INVENTION: Modified Forms of Hepatitis C NS3 Protease for

; TITLE OF INVENTION: Facilitating Inhibitor Screening and Structural Studies

; TITLE OF INVENTION: of Protease:Inhibitor Complexes

; FILE REFERENCE: DB17Sequences

; CURRENT APPLICATION NUMBER: US/09/965,594
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/115,271
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 588
; TYPE: DNA
; ORGANISM: Hepatitis C virus
US-09-965-594-13

Alignment Scores:
Pred. No.: 5,986-99 Length: 588
Score: 929.00 Matches: 183
Percent Similarity: 94.42% Conservatives: 3
Best Local Similarity: 92.89% Mismatches: 9
Query Match: 91.44% Indels: 2
DB: 10 Gaps: 1

US-09-965-594-22 (1-197) x US-09-965-594-13 (1-588)

Qy 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
Db 1 ATGAAAAAAGGATCGTGTATATCGTCGCCGTATAGTACTGAACGGT-----GCT 54
Qy 21 TyrAlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAsp 40
Db 55 TACGCTCAGCAGACTCGAGGTGAGGAGGTGAGCAAGAACTCCACAGACCGTGGTGC 114
Qy 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
Db 115 AAAAACCAGGTGAAGGTGAAGTTCAGATCGTTCCACCGTGTCTCAGACCTTCCTGGCT 174
Qy 61 ThrSerIleAsnGlyValLeuTyrValTyrHisGlyAlaGlyThrArgThrIleAla 80
Db 175 ACCTGCATCAACGGTGTGTGTCGACCGTTCACACCGTGTGGTACCGTACCATCGCT 234
Qy 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr 100
Db 235 TCCCGAAAGGTCCGGTTATCCAGATGTACACCAAGTTGACAAAGACCTGGTGGTGG 294
Qy 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
Db 295 CCGGCTCCGACGGTTCGGTTCCTGACCCCGTCACCTTCGGTTCCTCCGACCTGTAC 354
Qy 121 LeuValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySer 140
Db 355 CTGGTTACCCGTCACGCTGACGTTATCCCGGTTCGTCGTCGTCGTCGTCGTCGTC 414
Qy 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyProLeuLeuCys 160
Db 415 CTGGTGTCCCGCTCCGATCTCCATCCAGAAAGTTCCTCCGCTGGTGGTGGTGGT 474
Qy 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180
Db 475 CCGGCTGGTTCAGCGCTGTGGTATCTCCGTCGTCGTCGTCGTCGTCGTCGTCGTA 534
Qy 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
Db 535 GCTGTGTGACTTCATCCCGGTTCGAACTCCCTGGAAACCAACCATGCGTTCCTCC 585

RESULT 8

US-09-965-594-4
; Sequence 4, Application US/09965594
; Patent No. US20020106642A1
; GENERAL INFORMATION:
; APPLICANT: Wittekind, Michael
; APPLICANT: Weighelmer, Steven
; APPLICANT: Zhang, Yaquin
; APPLICANT: Goldfarb, Valentina
; TITLE OF INVENTION: Modified Forms of Hepatitis C NS3 Protease for
; Facilitating Inhibitor Screening and Structural Studies

; TITLE OF INVENTION: of Protease:Inhibitor Complexes
; FILE REFERENCE: DB17Sequences
; CURRENT APPLICATION NUMBER: US/09/965,594
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/115,271
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 588
; TYPE: DNA
; ORGANISM: Hepatitis C virus
US-09-965-594-4

Alignment Scores:
Pred. No.: 8,296-96 Length: 588
Score: 902.00 Matches: 181
Percent Similarity: 92.39% Conservatives: 1
Best Local Similarity: 91.88% Mismatches: 13
Query Match: 88.78% Indels: 2
DB: 10 Gaps: 1

US-09-965-594-22 (1-197) x US-09-965-594-4 (1-588)

Qy 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
Db 1 ATGAAAAAAGGATCGTGTATATCGTCGCCGTATAGTACTGAACGGT-----GCT 54
Qy 21 TyrAlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAsp 40
Db 55 TACGCTCAGCAGACTCGAGGTGCTGGGTGCATCATCCTCCCTGACCGCTGGTGC 114
Qy 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
Db 115 AAAAACCAGGTGAAGGTGAAGTTCAGATCGTTCCACCGTGTCTCAGACCTTCCTGGCT 174
Qy 61 ThrSerIleAsnGlyValLeuTyrValTyrHisGlyAlaGlyThrArgThrIleAla 80
Db 175 ACCTGCATCAACGGTGTGTGTCGACCGTTCACACCGTGTGGTACCGTACCATCGCT 234
Qy 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr 100
Db 235 TCCCGAAAGGTCCGGTTATCCAGATGTACACCAAGTTGACAAAGACCTGGTGGTGG 294
Qy 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
Db 295 CCGGCTCCGACGGTTCGGTTCCTGACCCCGTCACCTTCGGTTCCTCCGACCTGTAC 354
Qy 121 LeuValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySer 140
Db 355 CTGGTTACCCGTCACGCTGACGTTATCCCGGTTCGTCGTCGTCGTCGTCGTCGTC 414
Qy 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyProLeuLeuCys 160
Db 415 CTGGTGTCCCGCTCCGATCTCCATCCAGAAAGTTCCTCCGCTGGTGGTGGTGGT 474
Qy 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180
Db 475 CCGGCTGGTTCAGCGCTGTGGTATCTCCGTCGTCGTCGTCGTCGTCGTCGTCGTA 534
Qy 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
Db 535 GCTGTGTGACTTCATCCCGGTTCGAACTCCCTGGAAACCAACCATGCGTTCCTCC 585

RESULT 9

US-10-133-133A-6
; Sequence 6, Application US/10133133A
; Publication No. US20030114385A1
; GENERAL INFORMATION:
; APPLICANT: CATHERS, Brian
; APPLICANT: NEUTEBOON, Saskia
; APPLICANT: SHEPARD, Michael
; TITLE OF INVENTION: VIRAL ENZYME ACTIVATED PROTOTOXOPHORES

; TITLE OF INVENTION: AND USE OF SAME TO TREAT VIRAL INFECTIONS

```
; FILE REFERENCE: NB 2021.00
; CURRENT APPLICATION NUMBER: US/10/133,133A
; CURRENT FILING DATE: 2002-04-26
; PRIOR APPLICATION NUMBER: 60/286,983
; PRIOR FILING DATE: 2001-04-27
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 612
; TYPE: DNA
; ORGANISM: Hepatitis C Virus
US-10-133-133A-6

Alignment Scores:
Pred. No.:      2,02e-91      Length:      612
Score:          864.50      Matches:      174
Percent Similarity: 90.77%      Conservative: 3
Best Local Similarity: 89.23%      Mismatches:  15
Query Match:      85.09%      Indels:       3
DB:              14          Gaps:         1

US-09-965-594-22 (1-197) x US-10-133-133A-6 (1-612)
```

```
QY 5 GlySerValValIleValGlyArqIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
Db 19 GCTAGTGTGGTCATTTGGGTAGGATCATTTTCCGGTAGTGGTATGATCATCGCGGTAC 78
QY 22 AlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAspLys 41
Db 79 GCCCAGCAGACAGGGCCCTCCTAGGTGCATATCACCAGCCTAATCGCCGGGACAAA 138
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
Db 139 AACCAAGTGGAGGTGAGGTGCCAGATTGTGTCACTGTCTCCCAAAACCTTCTCTGGCAACG 198
QY 62 SerIleAsnGlyValLeuTTPThrValTyrHisGlyValGlyThrArgThrIleAlaSer 81
Db 199 TGCATCAATGGGTGTCTGGAGTGTCTACACGGGGCCGGACAGGACCATCGCGTCA 258
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTTPGln 101
Db 259 CCCAAGGTCTCTGTCATCCAGATGTATACCAATGTAGACCAAGACCTTGTGGCTGCGCC 318
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
Db 319 GCTTCGCAAGGTACCCGCTCATTCACACCTGTCACTTGGCGCTCCTCGGACCTTTACCTG 378
QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
Db 379 GTCACGAGGCACGCGATGTCATTCCTCGCGCGCGGGGTGATACGAGGGCAGCCTG 438
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
Db 439 CTGCGCGCGCGCGCATTTCTTACTTGAAGGTCTCTCGGGGGTCCGCTGTTGTGCGCC 498
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
Db 499 CGGGGGCAGCGCGTGGGCATATTTAGGGCCCGCGGTGTGCACCCGTGGAGTGGCTAAGCCG 558
QY 182 ValAspPheIleProValGlySerLeuGluThrThrMetArgSer 196
Db 559 GTGGACTTTATCCCTGTGGAGAACCTTAGAGACAACCATGAGGTCC 603
```

RESULT 10

```
US-09-742-659-3
; Sequence 3, Application US/09742659
; Patent No. US20010034019A1
; GENERAL INFORMATION:
; APPLICANT: Hong, Zhi
; APPLICANT: Butkiewicz, Nancy J.
; APPLICANT: Zhong, Weidong
; APPLICANT: Ingravallo, Paul
```

```
; APPLICANT: Wright-Minogue, Jacquelyn
; APPLICANT: Lau, Johnson Y.
; APPLICANT: Lemon, Stanley M.
; TITLE OF INVENTION: Chimeric HCV/GBV-B viruses
; FILE REFERENCE: ID01116
; CURRENT APPLICATION NUMBER: US/09/742,659
; CURRENT FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: US 60/171,469
; PRIOR FILING DATE: 1999-12-22
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 9646
; TYPE: DNA
; ORGANISM: Hepatitis C virus
US-09-742-659-3
```

```
Alignment Scores:
Pred. No.:      5,25e-88      Length:      9646
Score:          848.50      Matches:      168
Percent Similarity: 87.25%      Conservative: 10
Best Local Similarity: 82.35%      Mismatches:  17
Query Match:      83.51%      Indels:       9
DB:              9          Gaps:         1
```

```
US-09-965-594-22 (1-197) x US-09-742-659-3 (1-9646)
QY 3 LysLysGlySerValIleValGlyArgIleAsn----- 14
Db 335A COTAGGGCCAGGAGATACTGTTGGACACCGCCAGGAATGGTCTCCAAGGGGTGGAGG 3413
QY 15 ---LeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGlnGlyThrGlnLys 33
Db 3414 TTGCTGGCGCCCATCAGCGGTACGCCAGCAGACGAGAGCCCTCCTAGAGTGTATATC 3473
QY 34 ThrSerHisThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 53
Db 3474 ACCACTCTGACTGGCGGACAAAACCAAGTGGAGGTGAGGTCCAGATCGTGTCAACT 3533
QY 54 AlaThrGlnThrPheLeuAlaThrSerIleAsnGlyValLeuTyrThrValTyrHisGly 73
Db 3534 GCTACCCAAACCTTCTGCGCAACGTGCATCAATGGGGTATGCTGACTGTCTACACGGG 3593
QY 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValThrGlnMetTyrThrAsnVal 93
Db 3594 GCGGGAACGAGGACCATCGATCACCAGGGTCTCTCATCCAGATGTATACCAATGTG 3653
QY 94 AspLysAspLeuValGlyTTPGlnAlaProGlnGlySerArgSerLeuThrProCysThr 113
Db 3654 GACCAAGACCTTGTGGGTGGCGCGCTCTCTCAAGGTTCCCGCTCATTCACACCTGCACC 3713
QY 114 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 133
Db 3714 TCGGGCTCTCTCGGACCTTTTACCTGTTAGAGGACGCGCGACGTCATTTCCCGTGGCGCG 3773
QY 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 153
Db 3774 CGAGGTGATAGCAGGGTAGCTGCTTCGCCCGCGCCCATTTCTACTCTAAAGGCTCC 3833
QY 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 173
Db 3834 TCGGGGGGTCCGCTGTGTGTCGCCCGGAGACGCGCTATTTCAGGGCGCGGTG 3893
QY 174 SerThrArgGlyValAlaLysAlaValAspPheIleProValGlyLeuLeuGluThrThr 193
Db 3894 TGCACCCGTGGAGTGGCCAAAGCGGTGGACTTTATCTCTGTGGAGAACCTTAGACACACC 3953
QY 194 MetArgSerPro 197
Db 3954 ATGAGATCCCCC 3965
```

RESULT 11

```
US-09-238-076-1
```



```

Pred. No.: 5,25e-88 Length: 9646
Score: 848.50 Matches: 168
Percent Similarity: 87.25% Conservative: 10
Best Local Similarity: 82.35% Mismatches: 17
Query Match: 83.51% Indels: 9
DB: 11 Gaps: 1

US-09-965-594-22 (1-197) x US-09-995-937-1 (1-9646)

QY 3 LysLysGlySerValValIleValGlyArgIleAsn-----14
Db 3354 CGTAGGGCCAGAGATACGCTTGGCCAGCCAGCGAATGTCTCCAGGGGTGGAGG 3413
QY 15 ---LeuSerGlyAspThrAlaGlnGlnInThrArgGlyGluGlnGlyThrGlnLys 33
Db 3414 TTGCTGGCCCATCAGCGGTACGCCAGCAGCAGAGGCTCTCTAGGGTGTAATC 3473
QY 34 ThrSerHisThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 53
Db 3474 ACCAGCTGACTGGCCGGGACAAAACCAAGTGGAGGTGAGGTCCAGATCGTCAACT 3533
QY 54 AlaThrGlnThrPheLeuAlaThrSerIleAsnGlyValLeuTrpValThrValIleHisGly 73
Db 3534 GCTACCAACCTTCTGGCAACGTCATCAATGGGTATGCTGATCCAGATGTATACCAATGTC 3593
QY 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValThrGlnMetThrThrAsnVal 93
Db 3594 GCCGGAACGAGGACCATCGCATCCCAAGGTCCTGTCTCATCCAGATGTATACCAATGTG 3653
QY 94 AspLysAspLeuValGlyTrpClnAlaProGlnGlySerArgSerLeuThrProCysThr 113
Db 3654 GACCAAGACCTTGTGGGCTGGCCCGCTCTCAAGGTTCCCGCTCATTTACACCTGGACC 3713
QY 114 CysGlySerSerAspLeuValThrArgHisAlaAspValIleProValArgArg 133
Db 3714 TCGGCTCTCGGACCTTTACTGTGTCAGAGGACGCCGATGTCNTCCGTGCGCGG 3773
QY 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerThrLeuLysGlySer 153
Db 3774 CGAGGTGATAGCAGGGGTAGCTGCTTCGCCCGCCGCAATTTCTTCTTGAAGGCTCC 3833
QY 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 173
Db 3834 TCGGGGGGTCCGCTGTGTGGCCCGGACACGCGGTGGGCTTATTTCAGGCGCGGTG 3893

Alignment Scores: 5,25e-88 Length: 9646
Pred. No.: 848.50 Matches: 168
Score: 87.25% Conservative: 10
Best Local Similarity: 82.35% Mismatches: 17
Query Match: 83.51% Indels: 9
DB: 11 Gaps: 1

US-09-965-594-22 (1-197) x US-09-917-563-1 (1-9646)

QY 3 LysLysGlySerValValIleValGlyArgIleAsn-----14
Db 3354 CGTAGGGCCAGAGATACGCTTGGCCAGCCAGCGAATGTCTCCAGGGGTGGAGG 3413
QY 15 ---LeuSerGlyAspThrAlaGlnGlnInThrArgGlyGluGlnGlyThrGlnLys 33
Db 3414 TTGCTGGCCCATCAGCGGTACGCCAGCAGCAGAGGCTCTCTAGGGTGTAATC 3473
QY 34 ThrSerHisThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 53
Db 3474 ACCAGCTGACTGGCCGGGACAAAACCAAGTGGAGGTGAGGTCCAGATCGTCAACT 3533
QY 54 AlaThrGlnThrPheLeuAlaThrSerIleAsnGlyValLeuTrpValThrValIleHisGly 73
Db 3534 GCTACCAACCTTCTGGCAACGTCATCAATGGGTATGCTGATCCAGATGTATACCAATGTC 3593
QY 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValThrGlnMetThrThrAsnVal 93
Db 3594 GCCGGAACGAGGACCATCGCATCCCAAGGTCCTGTCTCATCCAGATGTATACCAATGTC 3653
QY 94 AspLysAspLeuValGlyTrpClnAlaProGlnGlySerArgSerLeuThrProCysThr 113
Db 3654 GACCAAGACCTTGTGGGCTGGCCCGCTCTCAAGGTTCCCGCTCATTTACACCTGGACC 3713
QY 114 CysGlySerSerAspLeuValThrArgHisAlaAspValIleProValArgArg 133
Db 3714 TCGGCTCTCGGACCTTTACTGTGTCAGAGGACGCCGATGTCNTCCGTGCGCGG 3773
QY 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerThrLeuLysGlySer 153
Db 3774 CGAGGTGATAGCAGGGGTAGCTGCTTCGCCCGCCGCAATTTCTTCTTGAAGGCTCC 3833
QY 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 173
Db 3834 TCGGGGGGTCCGCTGTGTGGCCCGGACACGCGGTGGGCTTATTTCAGGCGCGGTG 3893

RESULT 13
US-09-917-563-1
; Sequence 1, Application US/09917563
; Publication No. US20030073080A1
; GENERAL INFORMATION:
; APPLICANT: RICE, CHARLES et al.
; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
; VIRUS (HCY) AND USES THEREOF
;
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAFERKAMP, L.C.
; STREET: 7733 FORSYTH BLVD., SUITE 1400
; CITY: ST. LOUIS
; STATE: MO
; COUNTRY: USA
; ZIP: 63105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
Qy 174 SerThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 193
|||||
Db 3894 TCACCCGTTGGAGTGGCTAAGCGGTGGACITTTATCCCTGTGGAGAACCTAGAGACAC 3953
|||||
Qy 194 MetArgSerPro 197
|||||
Db 3954 ATGAGATCCCG 3965
|||||
RESULT 14
US-09-238-076-5
; Sequence 5, Application US/09238076
; Patent No. US20020102540A1
; GENERAL INFORMATION:
; APPLICANT: RICE, CHARLES et al.
; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
; TITLE OF INVENTION: VIRUS (HCV) AND USES THEREOF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; STREET: 7733 FORSYTH BLVD., SUITE 1400
; CITY: ST. LOUIS
; STATE: MO
; COUNTRY: USA
; ZIP: 63105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/238,076
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/034,756
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 6029-4831
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 314-727-5188
; TELEFAX: 314-727-6092
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12980 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-09-238-076-5
Alignment Scores:
Pred. No.: 7,72e-88 Length: 12980
Score: 848.50 Matches: 168
Percent Similarity: 87.25% Conservative: 10
Best Local Similarity: 82.35% Mismatches: 17
Query Match: 83.51% Indels: 9
DB: 10 Gaps: 1
US-09-965-594-22 (1-197) x US-09-238-076-5 (1-12980)
Qy 3 LysLysGlySerValIleValGlyArgIleAsn----- 14
|||||
Db 3354 CGTAGGGCCGAGGATGCTTGGCCAGCGGAGTGGCTCCAGGGGTGGAGG 3413
|||||
Qy 15 ---LeuSerGlyAspThrAlaIleGlnThrArgGlyGluGlnGlyThrGlnLys 33
|||||
Db 3414 TTGCTGGCGCCCATCAGCGGTACGCCAGCAGACGAGGCGCTCCTAGGTTATATC 3473
|||||
Qy 34 ThrSerHisThrGlyArgAspLysAsnGlnValGluGlyGluValSerThr 53
|||||
```

```
Db 3474 ACCAGCCTGACTGCGCGGGACAAAACCAAGTGGAGGTGAGGTCCAGATCGTGCTCAACT 3533
|||||
Qy 54 AlaThrGlnThrPheLeuAlaThrSerIleAsnGlyValLeuThrThrValThrHisGly 73
|||||
Db 3534 GCTACCCAAACCTTCCTGGCAAGTGCATCAATGGGTATGCTGSACTGCTACCAACGGG 3593
|||||
Qy 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValThrGlnMetTyrThrAsnVal 93
|||||
Db 3594 GCCGGAACGAGGACCATCGCATCACCACAGGTCTCTGTATCCAGATGTATACCAATGTG 3653
|||||
Qy 94 AspLysAspLeuValGlyTrpGlnAlaProGlnGlySerArgSerLeuThrProCysThr 113
|||||
Db 3654 GACCAAGACCTTGTGGGCTGGCCGCTCTCAAGGTTCGCGCTCATTCACACCTGCACC 3713
|||||
Qy 114 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 133
|||||
Db 3714 TGGGCTCCTCGGACCTTTACCTGGTCACGAGCAGCGCATGTCTATCCCGTCCCGCGG 3773
|||||
Qy 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 153
|||||
Db 3774 CGAGGTGATAGCAGGGGTAGCTGCTTTCGCCCGGCCCATTTCTTACTTTGAAAGGCTCC 3833
|||||
Qy 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 173
|||||
Db 3834 TCGGGGGTCCGCTGTTGTGCCCGCGGGACACGCGTGGCGCTATTCAGGGCGCGGTG 3893
|||||
Qy 174 SerThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 193
|||||
Db 3894 TGCACCGTGGAGTGGCTAAGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAC 3953
|||||
Qy 194 MetArgSerPro 197
|||||
Db 3954 ATGAGATCCCG 3965
|||||
RESULT 15
US-09-965-937-5
; Sequence 5, Application US/09995937
; Publication No. US20030028010A1
; GENERAL INFORMATION:
; APPLICANT: RICE, CHARLES et al.
; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAFERKAMP, L.C.
; STREET: 7733 FORSYTH BLVD., SUITE 1400
; CITY: ST. LOUIS
; STATE: MO
; COUNTRY: USA
; ZIP: 63105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/995,937
; FILING DATE: 28-No. US20030028010A1-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/034,756
; FILING DATE: 04-May-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 6029-4831
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 314-727-5188
; TELEFAX: 314-727-6092
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12980 base pairs
```

```
;
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-995-937-5

Alignment Scores:
Pred. No.: 7.72e-88 Length: 12980
Score: 848.50 Matches: 168
Percent Similarity: 87.25% Conservative: 10
Best Local Similarity: 82.35% Mismatches: 17
Query Match: 83.51% Indels: 9
DB: 11 Gaps: 1

US-09-965-594-22 (1-197) x US-09-995-937-5 (1-12980)

QY 3 LysLysGlySerValValIleValGlyArgIleAsn----- 14
Db 3354 CgTAGGGCCAGGAGACTGCTTGGGCCAGCCGCGGAATGGTCTCCAAGGGGTGGAGG 3413
QY 15 ---LeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGlnGlyThrGlnLys 33
Db 3414 TTGCTGGCGCCCATCAGCGCTACGCCAGCAGCAGAGAGGCTCTTAGGGTGTATAATC 3473
QY 34 ThrSerHisThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 53
Db 3474 ACCAGCTGACTGGCCGGGCAAAACCAAGTGGAGGGTGAGGTCCAGATCGTGTCAACT 3533
QY 54 AlaThrGlnThrPheLeuAlaThrSerIleAsnGlyValLeuThrThrValThrHisGly 73
Db 3534 GCTACCCAAACCTTCCTGGCAACGTGCATCAATGGGTATGCTGGACTGCTACCAAGGG 3593
QY 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValThrGlnMetTyrThrAsnVal 93
Db 3594 GCCGGAACGAGGACCATCGCATCACCCAAAGGCTCTGTCATCCAGATGTATACCAATGTG 3653
QY 94 AspLysAspLeuValGlyTyrGlnAlaProGlnGlySerArgSerLeuThrProCysThr 113
Db 3654 GACCAAGACCTTGTGGGCTGGCCGCTCCTCAAGGTTCCCGCTCATTTGACACCCCTGCACC 3713
QY 114 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 133
Db 3714 TCGGCTCTCTCGGACCTTTACTGTGTACAGGACGCGCGATGTCATTCCGTCGCGCGG 3773
QY 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 153
Db 3774 CGAGGTGATAGCAGGGGTAGCTGCTTTCCGCCCGGCCCATTTCTCTACTTGAAGGCTCC 3833
QY 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 173
Db 3834 TCGGGGGTCCGCTGTTGTCGCCCGGCGGACACGCCGTGGGCTATTACAGGCGCGCGTG 3893
QY 174 SerThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 193
Db 3894 TGCACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACC 3953
QY 194 MetArgSerPro 197
Db 3954 ATGATATCCCCG 3965
```

Search completed: August 31, 2003, 04:55:02
Job time : 189.482 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 30, 2003, 19:20:43 ; Search time 1910.31 Seconds
(without alignments)
2506.388 Million cell updates/sec

Title: US-09-965-594-22
Perfect score: 1016
Sequence: 1 MKKGSVVIVGRINLSGDTA.....VAKAVDFIPVESLETTMRSP 197
Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0
Searched: 22781392 seqs, 12152238056 residues
Total number of hits satisfying chosen parameters: 45562784
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Command line parameters: -DEV-rlp
-Q/cgn2_1/USPTO_spoal/US09965594/runat_29082003_151919_28322/app_query.fasta_1.2872
-DB-EST -QFMT-fastap -SUFFIX-rst -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS-bits -START=1 -END=1 -MATRIX-blosum62 -TRANS-human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE-LOCAL
-OUTFMT=ptc -NORM-ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09965594 -CGN_1_1_12630 -runat_29082003_151919_28322 -NCPU=6 -ICPU=3
-NO_MAP -LARGEQUERY -NEG-SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba:*
2: em_estin:*
3: em_estim:*
4: em_estov:*
5: em_estpl:*
6: em_estro:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pin:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gsl1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	104.5	10.3	1031	14	CB950999	CB950999 AGENCOURT
2	101	9.9	1199	13	BQ892487	BQ892487 AGENCOURT
C 3	100.5	9.9	1403	13	BQ926101	BQ926101 AGENCOURT
C 4	98	9.6	644	29	BX238988	BX238988 Danio rer
C 5	97.5	9.6	629	10	BG089727	BG089727 mab90e06.
C 6	96	9.4	1146	12	BM915803	BM915803 AGENCOURT
C 7	95.5	9.4	701	10	BF863244	BF863244 963042C02
C 8	95.5	9.4	772	29	CC406704	CC406704 PUHKD12TB
C 9	95.5	9.4	789	29	CC406705	CC406705 PUHKD12TD
C 10	95.5	9.4	984	10	BF304699	BF304699 601888252
C 11	95	9.4	701	29	B2342381	B2342381 ic83b11.b
C 12	94.5	9.3	1062	29	CNS0608H	AL410873 T3 end of
C 13	94.5	9.3	1440	12	BM467279	BM467279 AGENCOURT
C 14	93	9.2	528	12	BM402566	BM402566 SLA005F12
C 15	93	9.2	560	28	AQ538021	AQ538021 RPCI-11-3
C 16	93	9.2	1213	13	B0541777	B0541777 AGENCOURT
C 17	92.5	9.1	938	13	B0894657	B0894657 AGENCOURT
C 18	92.5	9.1	1026	29	B2567288	B2567288 pacs2-164
C 19	92	9.1	617	10	BE055938	BE055938 945020D03
C 20	92	9.1	886	9	AL571605	AL571605 AL571605
C 21	91.5	9.0	528	28	AQ620249	AQ620249 HS-5182-B
C 22	91.5	9.0	958	10	BG420860	BG420860 602452062
C 23	91.5	9.0	1035	10	BE888775	BE888775 601513689
C 24	91.5	9.0	1733	12	BM553374	BM553374 AGENCOURT
C 25	91	9.0	580	14	CA728398	CA728398 wdl1c.pk0
C 26	91	9.0	736	12	B1459445	B1459445 603200433
C 27	91	9.0	733	13	B0402910	B0402910 604139183
C 28	91	9.0	866	13	B0219343	B0219343 603758452
C 29	91	9.0	906	13	BX434207	BX434207 BX434207
C 30	91	9.0	917	12	B1911168	B1911168 603062737
C 31	90.5	8.9	586	12	B1329116	B1329116 602980282
C 32	90.5	8.9	670	29	B2552327	B2552327 pacs1-60
C 33	90.5	8.9	812	13	B0299264	B0299264 603608431
C 34	90.5	8.9	814	11	CNS09179	BX059665 Single re
C 35	90.5	8.9	817	13	B0240438	B0240438 603323077
C 36	90.5	8.9	824	13	B0396924	B0396924 603405002
C 37	90.5	8.9	878	13	B0365755	B0365755 603584734
C 38	90.5	8.9	920	13	B0593458	B0593458 AGENCOURT
C 39	90.5	8.9	1106	13	B0956626	B0956626 AGENCOURT
C 40	90.5	8.9	1141	11	AK080345	AK080345 Mus muscu
C 41	90	8.9	500	12	BM708007	BM708007 UI-E-C11-
C 42	90	8.9	569	12	BM825317	BM825317 K-EST0097
C 43	90	8.9	590	10	BE382750	BE382750 601298414
C 44	90	8.9	631	10	AW961059	AW961059 EST373026
C 45	90	8.9	658	12	BM830847	BM830847 K-EST0104

ALIGNMENTS

RESULT 1
CB950999
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE

CB950999 1031 bp mRNA linear EST 29-APR-2003
AGENCOURT_13445496 NIH_MGC_177 Mus musculus CDNA clone
IMAGE:30316162 5', mRNA sequence.
CB950999
CB950999.1 GI:30205777
EST.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 1031)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs@mail.nih.gov
 Tissue Procurement: Dr. Michael Brownstein
 cDNA Library Preparation: Michael Brownstein Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: NDCM107 row: b column: 11
 High quality sequence stop: 333.
 Location/Qualifiers

FEATURES
 source
 1..1031
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:30316162"
 /lab_host="DH10B (T1-phage-resistant)"
 /clone_lib="NIH_MGC_177"
 /note="Organ: liver; Vector: pDNR-LIB; Site_1: SfiI (ggccattatgcc); Site_2: SfiI (ggccattcgcc); cDNA made by oligo-dT priming and directionally cloned. 5' and 3' adaptors were used in cloning as follows:
 5'-AAGCAGTGGTATCAGCAGAGTGGCCATTCAGCGCGG-3' and 5'-ATTCTAGCCGACGGCGGACATG-DT(30)NN-3'. Full-length enriched library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5 kb size fraction. Library created in the laboratory of M. Brownstein (NIH, NIH). Note: this is a NIH_MGC Library."
 235 a 309 c 211 g 275 t 1 others

Alignment Scores:
 Pred. No.: 3.53 Length: 1031
 Score: 104.50 Matches: 51
 Percent Similarity: 41.10% Conservative: 16
 Best Local Similarity: 31.29% Mismatches: 62
 Query Match: 10.29% Indels: 35
 DB: 14 Gaps: 8

US-09-965-594-22 (1-197) x CB950999 (1-1031)

Qy 44 ValGluGlyValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThrSerIle 63
 Db 395 ATTCAGGCTATCCCAAAACAGAGTACATCGGCAGCTTTTCCT---CACTCATTT 451
 Qy 64 AsnGlyValLeuThrValThrHisGlyAlaGlyThrArgThrIleAlaSerProLys 83
 Db 452 TTGGGCACACTGGTCCGTGGGCACAT-----ATCATCGCCCTAA 493
 Qy 84 GlyProValThrGlnMetThrThrAsnValAspLysValGlyTrpGlnAlaPro 103
 Db 494 GGGCCCTTCACAAA-----ACACTTAACCT-CCTTGCCTGGCCTGGCATGTGGG 543
 Qy 104 Gln-----GlySerArgSerLeuThrProCysThrCysGlySerSerAsp 118
 Db 544 CAAAGACCGTTTGGCTTCCTGGCTTCCTGGCCCCCCCCCAATTTGGGACACGTCG 603
 Qy 119 LeuThrLeuValThrArgHisAlaAsp-ValIleProValArgArgGlyAspSerAr 138
 Db 604 -----ACCACCCATGGGCTGTGTGTTCCCGCTCTCCCGTGGGCAATACA 651
 Qy 138 gGlySerLeuLeuSerProArgProIleSerThrLeuLysGlySerSerGly----- 155
 Db 652 AAACNCCCTTAAACCGTCCCTCCCAACAATATTCTTCAAGCGTCTCTGATTCCCTAA 711
 Qy 156 -GlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerTh 175
 Db 712 GTCCCCCTTTGTTACCCAGACCAATTTGTGGGACACAGCGGCTCTTTTATCTTC 771

Qy 175 rArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMetAr 195
 Db 772 C-----CCCCCTCATGCTCTT---CCCACACGGCG 798
 Qy 195 gSerPro 197
 Db 799 AACACCC 805
 RESULT 2
 LOCUS B0892487 1199 bp mRNA linear EST 16-AUG-2002
 DEFINITION AGENCOURT_8417538 Lupski_sympathetic_trunk Homo sapiens cDNA clone
 IMAGE:6192708 5', mRNA sequence.
 ACCESSION B0892487
 VERSION B0892487
 KEYWORDS EST.
 SOURCE B0892487.1 GI:22284501
 ORGANISM Homo sapiens (human)

REFERENCE 1 (bases 1 to 1199)
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs@mail.nih.gov
 Tissue Procurement: Dr. James R. Lupski
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LRAM1359S row: c column: 13
 High quality sequence start: 57
 High quality sequence stop: 394.
 Location/Qualifiers

FEATURES
 source

1..1199
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6192708"
 /sex="male"
 /tissue_type="sympathetic trunk"
 /dev_stage="adult, 16 yr"
 /lab_host="DH10B"
 /clone_lib="Lupski_sympathetic_trunk"
 /note="Vector: pCMV-SPORT6 (Life Technologies); Site_1:
 NotI; Site_2: SalI; cDNA made by oligo-dT priming.
 Directionally cloned using the following adaptors:
 5'-TCGACCCACAGCGTCGG-3' and
 5'-CACTAGTCTCATGTCGCGCGCGCT(15)-3'. Size selected >
 1 kb for average insert length 1.9 kb. This is a primary
 library, non-amplified. Library constructed by Life
 Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
 College of Medicine); available through Life
 Technologies."

BASE COUNT 255 a 362 c 343 g 211 t 28 others
 ORIGIN

Alignment Scores:
 Pred. No.: 9.59 Length: 1199
 Score: 101.00 Matches: 50
 Percent Similarity: 34.74% Conservative: 24
 Best Local Similarity: 23.47% Mismatches: 80
 Query Match: 9.94% Indels: 59
 DB: 13 Gaps: 9

US-09-965-594-22 (1-197) x B0892487 (1-1199)

Qy 16 SerGlyAspThrAlaTyrAlaGlnThrArgGlyGluGlnGlyThrGlnLysThrSer 35
 Db 337 GCAGGAGAGAAACCTTACCCCAACAG-----AAGGCA 369

```

Qy 36 HisThrGlyArgAspLysAsnGlnValGluGlyValGlnIleValSerThr----- 53
Db 370 CATGGCGGAAATCCCGCTTCACAGACGAGTCTTCATGTTTCTGAAACATAACCG 429
Qy 54 AlaThrGlnThrPheLeu-----AlaThrSerIleAsnGlyValLeuTrpThr 69
Db 430 CCAGCCACTGCTTCATGTAATACCTTCCACCACACACAGCGGCACACGATGGGAI 489
Qy 70 ValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLysGlyProValThrGlnMet 89
Db 490 CCATTTTAAAGAGGTGCTCTTAAATCATGCGCCACGCGCGCTGATCTTCCA 549
Qy 90 TyrThrAsnValAspLysLeuValGlyTrpGlnAlaProGlnGlySerArgSerLeu 109
Db 550 TTTACCACATGTGACAGTGACTT-----CysGlySerSerAspLeuTyr 120
Qy 110 ThrProCysThr-----GlySerSerGlyGlyPyr 157
Db 577 GCTGCTGCACAGCACCCCATGACCATGTGGGCTTATGTGGAACGCGGAGCGGTT 636
Qy 121 -LeuValThr-----ArgHisAlaAspValIleProValArg----- 132
Db 637 ATTGGCCACTCCCTCTATATAAACACAGCCACGCTGTTCCATGGCGCGGCTGGTGT 696
Qy 133 -----ArgArgGlyAspSerArgGlySerLeuLeu----- 142
Db 697 TTGGCAGCGCAAGCGGGTGGGGCATGTAGGACTCGGGGGCGATCTCTGAAACCC 756
Qy 143 -SerProArgProIleSerTyrLeuLys-----GlySerSerGlyGlyPyr 157
Db 757 CCACCTCGGGCCACCGATGGCTTAAGCCTCCCTTACAGCCACCGCGCGGCGCC 816
Qy 157 oLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgG1 177
Db 817 CCTAACATCTCTACCTCCCTGGCGCGGGGAGAGCTGGGGGCATACGGGCTCAGG 876
Qy 177 yValAlaLysAlaValAspPheIleProValGluSer 189
Db 877 CGTTTAAAGCCCGCGCTTCGCGCGCGGAGCA 913

RESULT 3
B0926101/c
LOCUS B0926101 1403 bp mRNA linear EST 20-AUG-2002
DEFINITION AGENCOURT_8752655 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6335718
5', mRNA sequence.
ACCESSION B0926101
VERSION B0926101.1 GI:22341132
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC http://mgi.nci.nih.gov/
1 (bases 1 to 1403)
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,
Ph.D.
cDNA Library Preparation: ResGen, Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAH13798 row: j column: 07
High quality sequence stop: 101.
Location/Qualifiers
1..1403
/mol_type="mRNA"
/organism="Mus musculus"
source

```

```

/db_xref="taxon:10090"
/clone="IMAGE:6335718"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_130"
/note="Organ: oocytes; Vector: pCMV-SPORT6.1.cdb;
Site_1: EcoRV; Site_2: NotI; Cloned unidirectionally.
Primer: Oligo dT. Average insert size 1.95 kb.
Constructed by ResGen, Invitrogen Corp. Note: this is a
NIH_MGC Library."
BASE COUNT 297 a 521 c 237 g 345 t 3 others
ORIGIN
Alignment Scores:
Pred. No.: 13.2 Length: 1403
Score: 100.50 Matches: 56
Percent Similarity: 35.32% Conservative: 15
Best Local Similarity: 27.86% Mismatches: 74
Query Match: 9.89% Indels: 56
DB: 13 Gaps: 9
US-09-965-594-22 (1-197) x B0926101 (1-1403)
Qy 4 LysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAlaTyrAlaGln 23
Db 1381 AGAGGTGTGTCANCGTCAGGACAGGTC---GCCGCACATCGACGGTCGGCCAGAG 1325
Qy 24 GlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAspLysAsnGln 43
Db 1324 ACTTGTGGGGCGGCTTGGCGCATACCCGGGTGGATCGAGGTGAGCGCGCTTGAT 1265
Qy 44 ValGluGlyValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThrSerIle 63
Db 1264 ACAGAGGGGAAA----- 1253
Qy 64 AsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLys 83
Db 1252 CAGGGGGA---TGGTTATCACGGGCTGGGCGAGGTACT-----TCCCTAAA 1208
Qy 84 --GlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGlnAlaP 103
Db 1207 GCGGCGGTGGCGGAGTATATATACCGCGAGTGGCAAGCGCGGCGGTGGAACGTG 1148
Qy 103 roGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeuVal 123
Db 1147 ACCAA---CAGAGGCACTTACGCCCTCCCTCGTGGGCTGCGATAATAACAAATGT 1091
Qy 123 hrArgHisAlaAspValIleProValArgArgGlyAsp----- 136
Db 1090 CAGGCGCGGTGATGTGTTACTACCGCGAGACCGGCTCCACGCGGCTCTCTACAGA 1031
Qy 137 -----SerArgGlySerLeuLeuSerProArgProIle-SerTyrLeuLysGlySer 153
Db 1030 CGCCCGCTCCCGCGCAAC-----AGGTAATAATCATATCATATCGGCGGGAT 983
Qy 154 SerGly-----GlyProLeuLeu 159
Db 982 TTCGATTCGCGGAGAGCGCGGTGCGGGGCGCGCGGTGCGGGCGGTGAGG 923
Qy 160 CysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyVal 178
Db 922 CGCAGGAGAGGC-----GCCGTGTTTCGCGGCTGAGGACGAGCGCGGTG 875

RESULT 4
BX238988/c
LOCUS BX238988 644 bp DNA linear GSS 29-JAN-2003
DEFINITION Danio rerio genomic clone DREY-283113, genomic survey sequence.
ACCESSION BX238988
VERSION BX238988.1 GI:28161322
KEYWORDS GSS.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;

```

REFERENCE	LOCUS	DEFINITION
<p>Cypriniformes: Cyprinidae; Danio.</p> <p>1 (bases 1 to 644)</p> <p>Humphray,S.J., Huckle,E. and Durham,J.L.</p> <p>Direct Submission</p> <p>Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humphray@sanger.ac.uk unpublished</p> <p>This sequence was generated from the T7 end of BAC 283L13. 283L13 is part of the Daniokey BAC Library created by R. Plasterk and N.V. Keygene. Further details: http://www.sanger.ac.uk/Projects/D_danio/</p>	<p>1. 644</p> <p>/organism="Danio rerio"</p> <p>/mol_type="genomic DNA"</p> <p>/db_xref="taxon:7955"</p> <p>/clone="DKEY-283L13"</p> <p>/tissue_type="Testis"</p> <p>/note="vector pIndigoBAC-536"</p>	<p>129 a 212 c 176 g 127 t</p>
<p>BASE COUNT</p> <p>ORIGIN</p>	<p>Alignment Scores:</p> <p>Pred. No.: 8.45 Length: 644</p> <p>Score: 98.00 Matches: 48</p> <p>Percent Similarity: 41.95% Conservative: 25</p> <p>Best Local Similarity: 27.59% Mismatches: 84</p> <p>Query Match: 9.65% Indels: 17</p> <p>DB: 29 Gaps: 7</p>	<p>US-09-965-594-22 (1-197) x BX238988 (1-644)</p>
<p>Qy 18 AspThrAlaTyrAlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThr 37</p> <p>Db 556 GACCATGGCTATTACCATCATCGGAGGGGAGGAGCGAGTCGGGAATCCGGAGCGCCATCTC 497</p> <p>Qy 38 GlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThr 57</p> <p>Db 496 CTTGACGGTGTGACGACCGAGCGGAGCCCTCCCTGGCCGGTGCAGATGCCATTAGGGCC 437</p> <p>Qy 58 PheLeuAlaThrSerIleAsnGlyValLeu-TrpThrValTyrHisGlyAlaGlyThrAr 77</p> <p>Db 436 GAGACTGCATGGACACCGCGCGTGGCAGCGTGGACAGCATCTGCGAGTGGGAGCTC 377</p> <p>Qy 77 gThrIleAlaSerProLysGlyProValThrGlnMetTyr-----ThrAs 92</p> <p>Db 376 TCGAGGGCGCTGTCGCGCGTCACACAGTGCAGCAGCAGTGCATGTCGTGTGTCAGCAGC 317</p> <p>Qy 92 nValAspLysAspLeuValGlyTyrGlnAla---ProGlnGlySerArg-----Se 108</p> <p>Db 316 TCTTGGNAGGAC-----TGGAGATCCACACCGTTGGGGACAAAGCGGAGGAGTC 266</p> <p>Qy 108 rLeuThrProCysThrCysGlySerSerAspLeuTyrLeuValThrArgHisAlaaspVa 128</p> <p>Db 265 GTACACCACTGCCCTTTAAAAAAGGAGAAATTTATTTCATTACCATGGGAGGAGTTG 206</p> <p>Qy 128 lIleProValArgArgGlyAspSerArgGlySerLeuLeuSerProArgPro----- 146</p> <p>Db 205 TCGCACGAGGAGGCTTAGGAGGAGGAGGCGCCCTCCCGGTGCTCTCCCTCCCTACATG 146</p> <p>Qy 147 -IleSerTyrLeuLysGlySerSerGlyGlyProLeuLeu---CysProAlaGlyHisAl 165</p> <p>Db 145 TATTCT---TTAAGGCGCTTGGGAGGAGCCCTTTGCTGGATGCCAGTCCCTGCCCC 89</p> <p>Qy 165 aValGlyIlePheArgAlaValSerThrArgGlyVal 178</p> <p>Db 88 TGCCCTGTCTACCCGGCGGAGAAACCTTCGAGGAGCTTA 49</p>		
<p>RESULT 5</p> <p>BG089727/c</p> <p>LOCUS</p> <p>DEFINITION</p>	<p>629 bp mRNA linear, EST 26-JAN-2001</p> <p>mab90e06.x1 NCI_CGAP_Sp2 Mus musculus cDNA clone IMAGE:3977578 3'</p> <p>similar to SW:GRAD_MOUSE P11033 GRANZYME D PRECURSOR i, mRNA</p>	<p>sequence.</p> <p>BG089727 1 GI:12572290</p> <p>EST.</p> <p>SOURCE</p> <p>ORGANISM</p> <p>Mus musculus (house mouse)</p> <p>Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.</p> <p>1 (bases 1 to 629)</p> <p>NCI-CGAP Http://www.ncbi.nlm.nih.gov/ncicgap.</p> <p>National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index</p> <p>Unpublished</p> <p>Contact: Robert Strausberg, Ph.D.</p> <p>Email: cgaps-r@mail.nih.gov</p> <p>Tissue Procurement: David Segal Ph.D., Herbert Morse M.D.</p> <p>cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)</p> <p>DNA Sequencing by: Washington University Genome Sequencing Center</p> <p>Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov</p> <p>MGI:1477610</p> <p>Seq primer: -40UP from Gibco</p> <p>High quality sequence stop: 422.</p> <p>Location/Qualifiers</p> <p>1. 629</p> <p>/organism="Mus musculus"</p> <p>/mol_type="mRNA"</p> <p>/db_xref="taxon:10090"</p> <p>/clone="IMAGE:3977578"</p> <p>/tissue_type="NK cells (flow-sorted)"</p> <p>/lab_host="DH10B (T1-resistant)"</p> <p>/clone_lib="NCI_CGAP_Sp2"</p> <p>/notes="Organ: spleen; Vector: pCMV-SPORT6 (Life Technologies); mRNA made from flow-sorted NK cells, cDNA made by oligo-dT priming. Directionally cloned. Average insert size 1.5 kb. Primary library, non-amplified. cDNA Library Preparation: David B. Krizman, Ph.D."</p> <p>BASE COUNT 131 a 156 c 150 g 191 t 1 others</p> <p>ORIGIN</p> <p>Alignment Scores:</p> <p>Pred. No.: 9.19 Length: 629</p> <p>Score: 97.50 Matches: 48</p> <p>Percent Similarity: 39.23% Conservative: 23</p> <p>Best Local Similarity: 26.52% Mismatches: 51</p> <p>Query Match: 9.60% Indels: 59</p> <p>DB: 10 Gaps: 12</p> <p>US-09-965-594-22 (1-197) x BG089727 (1-629)</p> <p>Qy 36 HisThrGlyArgAspLysAsnGlnValGluGlyValGlnIleValSerThrAlaThr 55</p> <p>Db 620 CATCCGGTAAG-----GAAGGAGACACAGATCATCCCTGTGCA--- 579</p> <p>Qy 56 GlnThrPheLeuAlaThrSerIleAsnGlyValLeuTrpThrValTyrHisGly----- 73</p> <p>Db 578 AAAACATTTTCCCATCCAGATATAAATGCT-----ACTATCTTCCAGGTGACATC 538</p> <p>Qy 74 -----AlaGlyThrArgThrIleAlaSer 81</p> <p>Db 527 ATGCTGTTAAGCTGGAGAGTAAGCCCAAGAGAACTAAAGCTGTGAGACCCCTCAAGTTG 468</p> <p>Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101</p> <p>Db 467 CCCAGATCCAATGCCGGGTGAAGCCAGCAATGCTG---TGCAGTGTGGCTGGCTGG--- 414</p> <p>Qy 102 AlaProGlnGlySerArgSerLeu-----ThrProCysThrCysGlySerSerAspLeu 119</p> <p>Db 413 -----GGGTCAAGTCCATCAATGACACTAAAGCATCTGCCCGCTCGGAGAGTT 363</p>

```

QY 120 TyrLeuValThrArgHisAlaAspValIleProValArgArgArg----- 134
DB 362 CAACCTGGTCATCCAGGAGGAGGAGGAGTCAAAAACGTTTCCGATACTACACTGAGACC 303
QY 135 -----GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyr 149
DB 302 ACAGAGATTTCGTGGAGACTTGAAG---AAAATAAGACTCTC----- 261
QY 150 LeuLysGlySerSerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePhe 169
DB 260 TTCAGGGTGACTCGGGGGACCCCTTGTGTGTGAC---AACCAAGCATATGGACTTTC 204
QY 170 ArgAlaAla-----ValSerThrArgGlyValAlaLysAlaValAspPheIle 185
DB 203 GCCTATGCAAAAACGGAACAATCTCTTCAGGAATCTTCACAAAGTTGTGCACCTCCTG 144
QY 186 Pro 186
DB 143 CCG 141

RESULT 6
BM915803/c
LOCUS BM915803 1146 bp mRNA linear EST 12-MAR-2002
DEFINITION AGENCOURT_6639455 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5482056
5', mRNA sequence.
ACCESSION BM915803
VERSION BM915803.1 GI:19366182
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1146)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/PTP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2007 row: i column: 01
High quality sequence start: 6
High quality sequence stop: 256.
High quality Location/Qualifiers
FEATURES
source
1..1146
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5482056"
/tissue_type="melanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_41"
/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
BASE COUNT 169 a 492 c 344 g 141 t
ORIGIN

Alignment Scores:
Pred. No.: 28.5 Length: 1146
Score: 96.00 Matches: 46
Percent Similarity: 35.71% Conservative: 19
Best Local Similarity: 25.27% Mismatches: 60

```

```

Query Match: 9.45% Indels: 57
DB: 12 Gaps: 10
US-09-965-594-22 (1-197) x BM915803 (1-1146)
QY 22 AlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSer-HisThrGlyArgAspIly 41
DB 1098 GCCACAGCGGTCTCGCGAGCGAGGTCGCTCCGCTCTGCTACGTCGCGTGGGAG 1039
QY 41 sAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaIth 61
DB 1038 CTGAGGCGACAGAGT-----CTACGCGGTGGGTAGGGGA 1003
QY 61 rSerIleAsnGlyValLeuTrpThrValTyrHis-----GlyAlaGlyThr-- 76
DB 1002 CGCGCTGTGTGGATGTTGTG-----TATCACTCCCGCGCGGGGAGGTACGTG 949
QY 77 -----ArgThrIleAlaSerPro-----LysGlyProValThrGlnMetTy 90
DB 948 AGCGAGGGCGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 903
QY 90 rThrAsnValAspLysAspLeuValGlyTrpGlnAlaProGln-----GlySerAr 107
DB 902 -----CAGATGTCGGGTGGGAGCGCCGCTCGCGCGGTGGGCGCCAG 859
QY 107 gSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeuValThrArgHisAlaAs 127
DB 858 ACTTGCTTGTGCTGTTTCTGTGG-----834
QY 127 pValIleProValArgArgArgGlyAspSerArgGlySerLeuLeuSerProArgProIl 147
DB 833 -----CGGAGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCT 787
QY 147 eSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGl 167
DB 786 CGGGTACTACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 742
QY 167 yIlePheArgAlaAlaValSerThrArgGlyValAlaLysAlaValAspPhe---IlePr 186
DB 741 GGCCTTCGGCGTCTGTGTGTTCTTCGCGGTCTGCGCGGGGGGGGGGTTTCGCGTACC 682
QY 186 oVal 187
DB 681 TTG 678

RESULT 7
BF863244 701 bp mRNA linear EST 19-JAN-2001
LOCUS BF863244
DEFINITION 963042C02.xl C. reinhardtii CC-1690, Stress condition I, normalized
, Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION BF863244
VERSION BF863244.1 GI:12253388
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii
ORGANISM Chlamydomonas reinhardtii
REFERENCE 1 (bases 1 to 701)
AUTHORS Grossman,A., Davies,J., Federspiel,N., Harris,E., Hauser,C.,
Lefebvre,P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.
TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants: project phase 3
JOURNAL Unpublished
COMMENT Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu
FEATURES
source
1..701

```



```

Qy 178 ValAlaLysAlaValAspPhe-----ilePro 186
    |||:|||||
Db 244 -----GCCATTGACATCAACAGGAGAGCTTCACCATGAGCCATATGCAATTCCC 194

Qy 187 ValGluSerLeuGlu 191
    :|:|:|:|:|
Db 193 TTGGATACATTGGAG 179

RESULT 9
LOCUS CC406705 789 bp DNA linear GSS 19-MAY-2003
DEFINITION PUHKL12TD 2M_0.6_1.0_KB Zea mays genomic clone ZMBBTA469B24,
genomic survey sequence.
ACCESSION CC406705
VERSION CC406705.1 GI:30886795
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
TITLE Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick
JOURNAL A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.
COMMENT Maize Genomics Consortium
Other_GSSs: PUHKL12TB
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: Tf
Class: sheared ends.
FEATURES             Location/Qualifiers
     source           1..789
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMBBTA469B24"
                     /clone_lib="2M_0.6_1.0_KB"
                     /note="Vector: PCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
BASE COUNT          210 a 220 c 203 g 156 t
ORIGIN
Alignment Scores:
Pred. No.:          19.6          Length:          789
Score:              95.50         Matches:         49
Percent Similarity: 39.46%        Conservative:    24
Best Local Similarity: 26.49%      Mismatches:     67
Query Match:        9.40%         Indels:         46
DB:                 29           Gaps:           10

US-09-965-594-22 (1-197) x CC406705 (1-789)

Qy 29 GlnGlyThrGlnLysThrSerHisThrGlyArgAspLysAsnGlnVal---GluGlyGlu 47
    |||
Db 110 CAGTCTCGTTCAGCGCCTCACCCCGCGAGATGGCGGAGACGAGTAAACAGGGCTAT 169

Qy 48 ValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThrSerIleAsnGlyVal--- 66
    |||
Db 170 GTTAAATTGTGACG-----AGCCCTATGTACGAGGCCACCGCTGTCCCGAGGCTATTCT 223

Qy 67 LeuThrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLysGlyProVal 86
    :|:|:|:|:|
Db 224 ATCTGGAGG-----TCACAGACTTTGCTGACGATGAGTCAAGGGTG 265

Qy 87 ThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr-----GlnAlaPro 103
    |||
Db 266 ACACAGATGATGACCCAGG-----AAGACGAGGAGCCCGAGTGGTTCCCTACATCAATAA 319

```

```

Qy 104 GlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeu----- 119
    |||
Db 320 CTGGGATTAGAGGAGGAGCACCACCATGCGAGCTGCGGGTAGTCTCCTCAACGGTCAGGAGCTGC 379

Qy 120 -----TyrLeuValThrArgHisAlaAspValIleProValArgArgGlyAspSer 137
    |||
Db 380 TGGCCCTACTTGACACAGGGTTCAACACATAACTTCATCACTGCAGCGGCA-CAACAG 438

Qy 138 ArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyPro 157
    |||
Db 439 CTGGGGTTACTTTGAACCCACACAGGTCGCGCATGTCAAGGTGGCAATGGAGACCCA 498

Qy 158 LeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGly 177
    :|:|:|:|
Db 499 GTTTTCTGCCAG-----CGAGTAACTCGTCGCGCA----- 528

Qy 178 ValAlaLysAlaValAspPhe-----IlePro 186
    |||:|||||
Db 529 -----GCCATTGACATCAACAGGAGAGTTCACCATGAGGCATATGCAATTCCC 579

Qy 187 ValGluSerLeuGlu 191
    :|:|:|:|:|
Db 580 TTGGATACATTGGAG 594

RESULT 10
LOCUS BF304699/c 984 bp mRNA linear EST 21-NOV-2000
DEFINITION BF304699
ACCESSION BF304699
VERSION BF304699.1 GI:11251586
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgc.ncl.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCW1005 row: g column: 13
High quality sequence stop: 646.
FEATURES             Location/Qualifiers
     source           1..984
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4122276"
                     /tissue_type="rhabdomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_17"
                     /note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI;
Site_2: XhoI; CDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAGGAG(G). Size-selected
for average insert size 1.8Kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

```

BASE COUNT          133 a 329 c 351 g 171 t
ORIGIN
Alignment Scores:
Pred. No.:          26.1          Length:          984

```

RESULT 12	CMS06QHN	1062 bp	DNA	linear	GSS 05-JUL-2001
LOCUS	CMS06QHN				
DEFINITION	73 end of clone AW00AA006B03 of library AW00A from strain CLIB 89 of yarrowia lipolytica, genomic survey sequence.				
ACCESSION	AL410673				
VERSION	AL410673.1	GI:12179275			
KEYWORDS	GSS.				
SOURCE	Yarrowia lipolytica				
ORGANISM	Yarrowia lipolytica				
	Yarrowia lipolytica				
	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				

full-length clones and constructed by Life Technologies.

Note: this is a NIH_MGC Library." 355 t

BASE COUNT 330 a 442 c 308 g 355 t 5 others

ORIGIN

Alignment Scores:

Pred. No.: 54.2 Length: 1440
Score: 94.50 Matches: 54
Percent Similarity: 36.89% Conservative: 22
Best Local Similarity: 26.21% Mismatches: 95
Query Match: 9.30% Indels: 35
DB: 12 Gaps: 9

US-09-965-594-22 (1-197) x BM467279 (1-1440)

Qy 12 ArgIleAsnLeuSerGlyAlaThrAlaGlnGlnThrArgGlyClnGlnGlyThr 31
Db 753 CAATCACCATATCCGGAGATGTGCTCTCTGT-----TTTAAAGGCTCACACA 800
Qy 32 GlnLysThrSerHisThrGlyArgAspLysAsnGln-----ValGlu-GlyGluValG1 49
Db 801 CCGCGCACCAACTCTACTGCCGCCACACAAAAATATACTTTCTGGAGCGGAATATCTT 860
Qy 49 nIleValSerThrAlaThrGlnThrPheLeu---AlaThrSerIleAsnGlyValLeuTr 68
Db 861 CTTCCGCCCCAGAGCAAGATGTTCTTTTATTTAGAAACGGGAGCGGTGCTTC 920
Qy 68 pThrValThrHisGlyAlaGlyThrArgThrIleAlaSerProLysGly-----Pr 85
Db 921 ATTTTTTTCCTCGCAGGCGACCTCTCGAATCCCGAGCGCGCGGTTCCTCGCTCC 980
Qy 85 oValThrGlnMetTyThrAsnValAspLysLeuValGlyTrp----- 100
Db 981 TACATCCCGAGTATATAAT-----CCCGGTGGGTGGGACGTTTCT 1022
Qy 101 -----GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySer-- 116
Db 1023 ACACCCACACCGTGGCCCTCTCTATCTATCTGCTATCTATCTATCTATCTATCTATCT 1082
Qy 117 -----SerAspLeuTyLeuValThrArgHisAlaAspValIleProVa 131
Db 1083 CCACACCGCATTTTACTCCCGCTATATCTTNTATCTGNTGCGGCGGACGCCGCC 1142
Qy 131 lArgArgArgGlyAspSer-----ArgGlySerLeuLeuSerProArgProIleSerTy 149
Db 1143 TAGGGGTGGGGGGGGCCATTTTTCACGGTAGC---ACATCGCCCGCCCTCATTTT 1199
Qy 149 rLeuLysGlySerSerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePh 169
Db 1200 TTGGTGGGGGGGGGGGGCGCCACCCCTCACCCCTCGTGGGGGAGTCGTCTCTCTC 1259
Qy 169 eArg-AlaAlaValSerThrArgGlyValAlaLysAlaValAspPheIleProValGlu 189
Db 1260 CTCTAGCACACTTCCATGAGCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1319
Qy 189 erLeuGluThr 193
Db 1320 TTCTCGTCCACCACA 1333

RESULT 14
BM402566 528 bp mRNA linear EST 01-JUL-2002
LOCUS
DEFINITION BM402566.1 34513 An expressed sequence tag (EST) collection from the
resurrection plant Selaginella lepidophylla 5, mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Selaginella lepidophylla
Selaginella lepidophylla
Eukaryote: Viridiplantae; Streptophyta: Embryophyta: Tracheophyta;
Lycopodiophyta; Isoetopsida; Selaginellales; Selaginellaceae;
Selaginella.

REFERENCE

1 (bases 1 to 528)

Iturriaga,G. and Cushman,J.C.

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished

Contact: Cushman JC

Department of Biochemistry

University of Nevada

MS200, Reno, NV 89557-0014, USA

Tel: 775-784-1918

Fax: 775-784-1650

Email: jcushman@unr.edu

PCR PRIMERS

FORWARD: T3 20mer

BACKWARD: T3 21mer

Plate: 005 row: F column: 12

Seq primer: T3 20mer

High quality sequence stop: 528.

Location/Qualifiers

1..528

FEATURES

source

/organism="Selaginella lepidophylla"

/mol_type="mRNA"

/db_xref="taxon:59777"

/clone="SLA005F12"

/tissue_type="microphyll fronds undergoing desiccation for

2.5 h"

/dev_stage="adult"

/clone_lib="An expressed sequence tag (EST) collection

from the resurrection plant Selaginella lepidophylla"

/note="Vector: Lambda Uni-zap XR, Bluescript SK-; Site_1:

EcoRI; Site_2: XhoI; Library construction was performed

according to manufacture's (Stratagene, Inc.) recommended

protocol for the Lambda Uni-zapXR vector and cDNA synthesis

kit."

BASE COUNT 129 a 125 c 137 g 137 t

ORIGIN

Alignment Scores:

Pred. No.: 20.5 Length: 528

Score: 93.00 Matches: 37

Percent Similarity: 42.98% Conservative: 15

Best Local Similarity: 30.58% Mismatches: 43

Query Match: 9.15% Indels: 26

DB: 12 Gaps: 4

US-09-965-594-22 (1-197) x BM402566 (1-528)

Qy 94 AspLysAspLeuValGlyTrpGlnAlaProGlnGlySerArgSerLeuThrProCysThr 113
Db 53 GACAAGGATGAGCGGTGCTGAGATCGATCGTCAAGCACAGATCTCAGCCCAATACCC 112
Qy 114 CysGlySerSerAspLeuTyLeuVal----- 122
Db 113 CTTGGAAGTTCGTCGATCTGCTTGTGGCCGAAGGTGATGCTATCGGTAATCTCTTT 172
Qy 123 -----ThrArgHisAlaAspValIleProValArgAlaGlyArgGlyAspSerArg 138
Db 173 GGATTGGATCATACGCTGACACAGCGGTCTATCGTCTTCGAAGGGAGATTACT--- 229
Qy 139 GlySerLeuLeuSerProArgProIleSerTyLeu----- 150
Db 230 ---TCAGCGGCTAATGGTCGTCCTCAAGACGTCGATCCAGACAGATGCCGCTATTAA 286
Qy 151 LysGlySerSerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArg 170
Db 287 CCGTGGAAACAGCGGGGGTCCGCTATTTCGAAATTTGATAGGCATCAACACT 346
Qy 171 AlaAlaValSerThrArgGlyValAlaLysAlaValAspPhe---IleProValGluSer 189
Db 347 GCTATATATTTCTCGCTGGCGCTTCATCAGCGGTGGGCTTTTCCATTCAGTTGACACG 406
Qy 190 Leu 190

```

Db          407 GTT 409
RESULT 15
AQ538021
LOCUS
DEFINITION   560 bp DNA linear GSS 18-MAY-1999
              RPCI-11-32014.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-32014,
              genomic survey sequence.
ACCESSION   AQ538021
VERSION     AQ538021.1 GI:4849711
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 560)
AUTHORS     Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
              ,J.C.
TITLE       Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
              Map Building
JOURNAL
COMMENT      Other_GSSs: RPCI-11-32014.TV
              Contact: Shaying Zhao, William Nierman, Mark Adams
              Department of Eukaryotic Genomics
              The Institute for Genomic Research
              9712 Medical Center Dr., Rockville, MD 20850
              Tel: 301 838 0200
              Fax: 301 838 0208
              Email: hbe@tigr.org
              Clones are derived from the human BAC library RPCI-11. For BAC
              library availability, please contact Pieter de Jong
              (pieter@dejong.med.buffalo.edu). Clones may be purchased from
              BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
              Research Genet cs (info@resgen.com). BAC end search page:
              http://www.tigr.org/tldb/hungen/bac_end_search/bac_end_search.html.
              Seq primer: SP6
              Class: BAC ends.
FEATURES             Location/Qualifiers
     source           1..560
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7622691"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-32014"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /clone_lib="RPCI-11"
                     /note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT        130 a 125 c 181 g 124 t
ORIGIN
Alignment Scores:
Pred. No.:        22.1      Length:      560
Score:            91.00     Matches:     44
Percent Similarity: 37.93%   Conservative: 11
Best Local Similarity: 30.34% Mismatches:   54
Query Match:      9.15%     Indels:      36
DB:               28        Gaps:         6

US-09-965-594-22 (1-197) x AQ538021 (1-560)

QY          36 HlsThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThr 55
            ||| |||||
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db          53 CACCCTGGAGGGGCTATGAGATCATAGAGGGGCTCTGCCAGTCTCTGTCATCCACTGCC 112
            |||
QY          56 GlnThrPheLeuAlaThrSerIleAsnGlyValLeuTrpThrValTyHisGlyAlaCly 75
            |||
            ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db          113 CAAGAG-----AATTGCAAAATGTGG-----CACGGATTGTGGA 145
            |||
QY          76 ThrArgThrIleAlaSerProLysGlyProValThrGlnMetTyThrAsp-ValAspLys 95
            |||
            ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db          146 CCCAGGATCCAGCTGGCTCTAAAGGG-----AAAACAGTGTCTCACTGTGTCTGCC 196

```

```

QY          95 sAspLeuValGlyTrpGlnAlaProGlnGlySer-----ArgSerLe 109
            ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db          197 AGAGCTGTGTGCTGTCGCAACACCCCGAGGAACTTGTGTCATGGGGAGGAGTCTGT 256
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY          109 uThrProCysThrCysGlySerSerAspLeuTyLeuValThrArgHisAlaAspVal 129
            : |||
            |||
Db          257 ACATAGATGT-----CTGGAGCAGAT 277
            |||
QY          129 eProValArgArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTy 149
            | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db          278 ATGACGAGCCATGGAGGAGAGAGAGCCCTTCTCCCTGGCCCGCCCTGGACCATA 337
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY          149 rLeuLysGlySer-SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIleP 169
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db          338 TCTGAAAGCTGTGAAGTGGAGGC-----TGC CGCGGGGGAGCTGATGCAGCTGTTTC 388
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY          169 heArgAlaAla 172
            |||
Db          389 TCACAGGGGCT 399

```

Search completed: August 31, 2003, 04:27:50
Job time : 1914.31 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 17:42:58 ; Search time 2.49162 Seconds
(without alignments)
700.745 Million cell updates/sec

Title: US-09-965-594-26
Perfect score: 49
Sequence: 1 GSWIVGRIVL 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues
Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49	100.0	11	21	AA15227 Hepatitis C virus
2	49	100.0	12	23	AA48242 Hepatitis C virus
3	49	100.0	13	22	AA96864 Hepatitis C virus
4	49	100.0	23	19	AAW40552 Synthetic nonstruc
5	49	100.0	23	22	AAW52536 Peptide KKS4a use
6	49	100.0	23	22	AA664158 KKS4a peptide. S
7	49	100.0	23	22	AA67391 KKS4a peptide. S
8	49	100.0	23	22	AA66371 Hepatitis C virus
9	49	100.0	23	23	ABG32508 Peptide KKS4a for

10	49	100.0	23	23	ABG31914 KKS4a peptide. S
11	49	100.0	23	23	AAE18687 NS4A peptide used
12	49	100.0	23	23	AAU76376 Hepatitis C virus
13	49	100.0	23	24	ABG72264 Hepatitis C virus
14	49	100.0	195	21	AA15212 Hepatitis C virus
15	49	100.0	195	21	AA15220 Hepatitis C virus
16	49	100.0	200	13	AA29846 HCV NS2-NS4 peptid
17	48	98.0	12	21	AA144731 Hepatitis C virus
18	48	98.0	13	22	AA96862 Hepatitis C virus
19	48	98.0	14	18	AAW13792 Hepatitis C virus
20	48	98.0	14	22	AAW4387 NS3 protease activ
21	48	98.0	14	22	AAW92337 Virus related pept
22	48	98.0	16	21	AAW54448 Peptide 4A4 used t
23	48	98.0	16	22	AA96851 Hepatitis C virus
24	48	98.0	16	22	AA96853 Hepatitis C virus
25	48	98.0	16	22	AA96854 Hepatitis C virus
26	48	98.0	17	21	AAW9552 Hepatitis C virus
27	48	98.0	17	21	AAW83773 HCV NS3A cofactor
28	48	98.0	17	21	AAW83775 HCV NS3A cofactor
29	48	98.0	17	22	AAW97114 Hepatitis C virus
30	48	98.0	18	23	ABW05367 NS4a peptide. Hep
31	48	98.0	23	20	AAW15763 Substrate peptide
32	48	98.0	23	21	AAW23810 Synthetic peptide
33	48	98.0	23	22	AA96855 Hepatitis C virus
34	48	98.0	23	22	AAW92336 Virus related pept
35	48	98.0	28	19	AAW37386 Hepatitis C virus
36	48	98.0	32	22	AA96856 Hepatitis C virus
37	48	98.0	34	16	AAW2856 NS3 serine proteas
38	48	98.0	36	19	AAW50782 Peptide used in im
39	48	98.0	54	16	AAW2855 NS4A protein. Hep
40	48	98.0	54	19	AAW37808 Nonstructural doma
41	48	98.0	54	20	AAW17898 Native HCV NS4A pe
42	48	98.0	63	15	AAW49651 HCV peptide C14-1.
43	48	98.0	86	18	AAW09051 Hepatitis C virus
44	48	98.0	87	15	AAW49652 HCV peptide C14-1.
45	48	98.0	87	17	AAW95545 HCV II chimeric ep

ALIGNMENTS

RESULT 1
AAB15227
ID AAB15227 standard; protein; 11 AA.
XX
AC AAB15227;
XX
DT 19-DEC-2000 (first entry)
XX
DE Hepatitis C virus NS4A residues 21-31.
XX
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
liver failure; liver cancer.
XX
OS Hepatitis C virus.
XX
PN WO200040707-A1.
XX
PD 13-JUL-2000.
XX
PF 06-JAN-2000; 2000WO-US00345.
XX
PR 08-JAN-1999; 99US-0115271.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX
DR WFI; 2000-465976/40.
XX
PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
amino acid, useful for screening inhibitors that may treat hepatitis C

PT -
 XX
 PS Claim 15; Page 22; 66pp; English.
 XX
 CC The present sequence comprises residues 21-31 of the Hepatitis C virus
 CC (HCV) NS4A protease enzyme. It was used to create a number of fusion
 CC proteins also containing the HCV NS3 protease. The NS4A and NS3 proteins
 CC are both essential for the replication of the virus, acting to cleave its
 CC replicative proteins from the polyprotein produced from the HCV genome.
 CC Inhibitors of the two proteins should be effective as antiviral
 CC treatments of HCV infection. This is useful as HCV can lead to chronic
 CC liver disease such as cirrhosis, liver failure and liver cancer. The
 CC present invention concerns a number of NS3 mutants and NS3-NS4A fusion
 CC proteins which can be used to identify inhibitors of this type, as well
 CC as enabling structural studies of the protease and protease:inhibitor
 CC complexes.
 XX
 SQ Sequence 11 AA;
 Query Match 100.0%; Score 49; DB 21; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.025;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GSVWIVGRIVL 11
 Db | | | | | | | | | |
 1 GSVWIVGRIVL 11
 RESULT 2
 AAM48242
 ID AAM48242 standard; Peptide; 12 AA.
 XX
 AC AAM48242;
 XX
 DT 21-MAR-2002 (first entry)
 XX
 DE Hepatitis C virus NS4A cofactor peptide.
 XX
 KW Pestivirus; Npro; protease; NS3; screening; NS4A.
 XX
 OS Hepatitis C virus.
 XX
 PN US6326137-B1.
 XX
 PD 04-DEC-2001.
 XX
 PF 25-JUN-1999; 99US-0344456.
 XX
 PR 25-JUN-1999; 99US-0344456.
 XX
 PA (SCHE) SCHERING CORP.
 XX
 PI Hong Z, Lai VCH, Lau JYN;
 XX
 DR WPI; 2002-121103/16.
 XX
 PT Nucleic acid construct encoding chimeric Hepatitis C virus (HCV)
 PT pestivirus genome where the Npro protease gene is replaced with NS3
 PT protease gene, useful for in vivo screening of compounds which inhibit
 PT HCV infection -
 XX
 PS Example 1; Fig 2; 20pp; English.
 XX
 CC The present invention relates to a nucleic acid construct encoding a
 CC chimeric Hepatitis C virus (HCV)-pestivirus genome. The construct
 CC comprises a pestivirus genome where a Npro pestivirus protease gene is
 CC replaced with a gene encoding a functional HCV NS3 protease. Furthermore,
 CC each junction site recognised by the Npro protease is replaced with a
 CC junction site recognised by the HCV NS3 protease. The construct is useful
 CC for screening compounds that inhibit HCV in vivo by inhibiting HCV
 CC protease, where screening may be in cell culture or in an animal model.
 CC The present sequence is a peptide from the tethered NS4A cofactor
 CC sequence from HCV NS3, which was used to illustrate the present

CC invention.
 XX
 SQ Sequence 12 AA;
 Query Match 100.0%; Score 49; DB 23; Length 12;
 Best Local Similarity 100.0%; Pred. No. 0.027;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GSVWIVGRIVL 11
 Db | | | | | | | | | |
 1 GSVWIVGRIVL 11
 RESULT 3
 AAB96864
 ID AAB96864 standard; peptide; 13 AA.
 XX
 AC AAB96864;
 XX
 DT 06-JUL-2001 (first entry)
 XX
 DE Hepatitis C virus NS2/3 cleavage inhibitory peptide SEQ ID NO: 22.
 XX
 KW Hepatitis C virus; HCV; NS2; NS3; inhibitory peptide; cleavage;
 KW replication inhibition; chimpanzee; human; infection; gene therapy.
 XX
 OS Hepatitis C virus.
 XX
 PN WO200116379-A1.
 XX
 PD 08-MAR-2001.
 XX
 PF 25-AUG-2000; 2000WO-US23444.
 XX
 PR 30-AUG-1999; 99US-0151395.
 XX
 PA (MERI) MERCK & CO INC.
 XX
 PI Darke PL, Jacobs AR, Kuo LC;
 XX
 DR WPI; 2001-343059/36.
 XX
 PT Inhibiting hepatitis C virus (HCV) replication in HCV infected cell, or
 PT in a patient or treating a patient for HCV infection comprises
 PT inhibiting autocleavage of NS2/3 -
 XX
 PS Disclosure; Page 46; 50pp; English.
 XX
 CC The present invention describes methods and compositions capable of
 CC preventing the replication of hepatitis C virus (HCV), involving
 CC administering a compound which inhibits NS2/3 autocleavage. Also provided
 CC are peptides capable of inhibiting this cleavage step, of which this
 CC sequence is an example. These are useful in the treatment of HCV
 CC infection in humans and chimpanzees, and in research applications, for
 CC example in studying the stabilisation of NS2/3, the effects of NS2/3 on
 CC HCV polyprotein processing and the effects of inhibiting NS2/3
 CC autocleavage.
 XX
 SQ Sequence 13 AA;
 Query Match 100.0%; Score 49; DB 22; Length 13;
 Best Local Similarity 100.0%; Pred. No. 0.03;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GSVWIVGRIVL 11
 Db | | | | | | | | | |
 1 GSVWIVGRIVL 11
 RESULT 4
 AAM40552
 ID AAM40552 standard; peptide; 23 AA.
 XX

```

AC AAW40552;
XX
XX
DT 20-NOV-1998 (first entry)
XX
XX Synthetic nonstructural peptide SNS4A.
DE
XX Synthetic Hepatitis C nonstructural protein; SNS4A peptide;
KW cofactor; NS3 protease.
KW
XX
XX Synthetic.
OS
OS Hepatitis C virus.
XX
XX WO9811134-A1.
XX
XX 19-MAR-1998.
XX
XX 12-SEP-1997; 97WO-US16182.
XX
XX 18-OCT-1996; 96US-0731336.
PR
PR 12-SEP-1996; 96US-0025274.
XX
XX (VERT-) VERTEX PHARM INC.
PA
XX
XX Fox T, Kim JL, Lin C, Morgenstern KA, Thomson JA;
PI
XX WPI; 1998-250953/22.
XX
XX New hepatitis C virus crystal compositions - comprising a HCV
PT NS3-like polypeptide complexed with a NS4A-like polypeptide, used
PT particularly for drug design
XX
XX Claim 4; Page 30; 97pp; English.
XX
XX This is the amino acid sequence of the novel SNS4A (synthetic
CC Hepatitis C nonstructural protein 4A) peptide. It acts as a cofactor
CC for the NS3 protease in order to achieve proteolytic processing of
CC Hepatitis C virus (HCV) nonstructural proteins. It is used in the
CC method of the invention as part of a device which can be used to
CC provide information for the design of drugs for the treatment of HCV
CC infection. They can also be used for determining the 3-dimensional
CC structure of molecules or molecular complexes which contain at least
CC some structurally similar features to a HCV NS3 serine protease domain.
XX
XX
XX Sequence 23 AA;
SQ
Query Match 100.0%; Score 49; DB 22; Length 23;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSVVIVGRIVL 11
Db 3 GSVVIVGRIVL 13

RESULT 6
AAG64158
ID AAG64158 standard; peptide; 23 AA.
XX
AC AAG64158;
XX
DT 19-OCT-2001 (first entry)
XX
DE KKSNS4a peptide.
XX
KW Hepatitis C virus; HCV; NS3 protease; alpha-ketoamide inhibitor;
KW virucide; hepatotropic; antiinflammatory; viral infection; KKSNS4a.
XX
OS Synthetic.
XX
XX WO200140262-A1.
XX
PD 07-JUN-2001.
XX
PF 01-DEC-2000; 2000WO-US32677.
XX
XX 03-DEC-1999; 99US-0168998.
XX
XX (DUPO ) DU PONT PHARM CO.
XX
XX PI Han W;
XX
XX WPI; 2001-464936/50.
XX
XX New ketoamide derivatives useful for treating infections e.g. hepatitis
PT C virus -
XX
XX Disclosure; Page 195; 282pp; English.
XX
XX The invention relates to novel ketoamide and ketoester derivatives
CC for use as inhibitors of hepatitis C virus (HCV) NS3 protease inhibitors.
CC The compounds are useful for treating viral infections e.g. hepatitis C
CC virus. The present sequence was used in an experiment measuring the
CC effect of an inhibitor on the rate of hydrolysis of an ester substrate.
XX
XX

```

SQ Sequence 23 AA:
 Query Match 100.0%; Score 49; DB 22; Length 23;
 Best Local Similarity 100.0%; Pred. No. 0.055;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSVWIVGRIVL 11
 |||||
 DB 3 GSVWIVGRIVL 13

RESULT 7
 AAB67391
 ID AAB67391 standard; peptide: 23 AA.
 AC AAB67391;
 XX
 XX 26-APR-2001 (first entry)
 DT
 XX
 XX KKN54a peptide.
 DE
 XX Lactam; hepatitis C virus; HCV; NS3 protease.
 KW
 XX
 XX Synthetic.
 OS
 XX
 XX WO200107407-A1.
 PN
 XX
 XX 01-FEB-2001.
 PD
 XX
 XX 26-JUL-2000; 2000WO-US20189.
 PF
 XX
 XX 26-JUL-1999; 99US-0145631.
 PR
 XX
 XX (DUPO) DU PONT PHARM CO.
 PA
 XX Priestley ES, Decicco CP;
 PI
 XX
 XX WPI; 2001-159696/16.
 DR
 XX
 XX New lactam derivatives are hepatitis C virus NS3 protease inhibitors
 PT useful for treating HCV infections
 PS
 XX Example 26; Page 100; 130pp; English.
 XX The present invention relates to Lactam derivatives. These derivatives
 CC may be used for treating hepatitis C virus (HCV) infection. They can
 CC also be used for inhibiting HCV in a body fluid sample and as a
 CC standard or reagent in a test or assay for determining the ability
 CC of a potential pharmaceutical to inhibit HCV NS3 protease and/or HCV
 CC growth.
 XX
 XX
 SQ Sequence 23 AA:
 Query Match 100.0%; Score 49; DB 22; Length 23;
 Best Local Similarity 100.0%; Pred. No. 0.055;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSVWIVGRIVL 11
 |||||
 DB 3 GSVWIVGRIVL 13

RESULT 8
 AAB66371
 ID AAB66371 standard; peptide: 23 AA.
 XX
 XX AAB66371;
 XX
 XX 09-APR-2001 (first entry)
 DT
 XX
 XX Hepatitis C virus protease inhibitor related peptide #1.
 DE
 XX Hepatitis C virus; protease; boronic acid; inhibitor; liver cirrhosis;
 KW

KW liver cancer; NS3; antiviral agent.
 XX
 XX Unidentified.
 XX
 XX WO200102424-A2.
 PN
 XX 11-JAN-2001.
 PD
 XX 07-JUL-2000; 2000WO-US18655.
 PF
 XX 07-JUL-1999; 99US-0142561.
 PR
 XX (DUPO) DU PONT PHARM CO.
 PA
 XX Kettner CA, Jagannathan S, Forsyth TP;
 PI
 XX WPI; 2001-103001/11.
 DR
 XX
 XX New boronic acid derivatives, optionally containing peptides, used to
 PT treat hepatitis C infections, are hepatitis C viral protease inhibitors
 PT
 XX Example 60; Page 208; 258pp; English.
 PS
 XX The present invention provides a number of boronic acid derivatives which
 CC act as inhibitors of the hepatitis C virus NS3 protease enzyme. They can
 CC be used to treat infection by the virus, which can cause liver cirrhosis
 CC and liver cancer.
 XX
 XX
 SQ Sequence 23 AA:
 Query Match 100.0%; Score 49; DB 22; Length 23;
 Best Local Similarity 100.0%; Pred. No. 0.055;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSVWIVGRIVL 11
 |||||
 DB 3 GSVWIVGRIVL 13

RESULT 9
 ABG32508
 ID ABG32508 standard; peptide: 23 AA.
 XX
 XX ABG32508;
 AC
 XX 15-NOV-2002 (first entry)
 DT
 XX Peptide KKN54a for HCV NS3 protease kinetic assay.
 DE
 XX
 XX NS3; HCV; protease; HCV infection; hepatitis; cirrhosis; liver cancer;
 KW pyrimidinone; serine protease inhibitor; virucide; hepatotropic;
 XX antiinflammatory; blood plasma processing; KKN54a.
 XX Synthetic.
 OS
 XX WO200248116-A2.
 PN
 XX 20-JUN-2002.
 PD
 XX 12-DEC-2001; 2001WO-US47911.
 PF
 XX 13-DEC-2000; 2000US-255290P.
 PR
 XX (BRIM) BRISTOL-MYERS SQUIBB PHARMA CO.
 PA
 XX Glunz PW, Douty BD, Han W;
 PI
 XX WPI; 2002-627251/67.
 DR
 XX
 XX New pyrimidinones useful as serine protease inhibitors in the treatment
 PT of e.g. viral infection
 PT
 XX

Example 140; Page 192; 270pp; English.

The invention relates to pyrimidinones of a formula given in the claims of the specification, their stereoisomers, salts and prodrugs. In assays, the pyrimidinone compounds inhibited Hepatitis C virus (HCV) NS3 protease with IC₅₀ values of less than 100 micro M. The compounds are useful for treating viral infection e.g. HCV infection (the causative agent of acute hepatitis and associated with cirrhosis and liver cancer) and as a reagent used as inhibitors of HCV protease in the processing of blood plasma for diagnostic and other commercial purposes. The present sequence is a peptide, KNNS4a, used in an NS3 kinetic assay.

Sequence 23 AA;

Query Match 100.0%; Score 49; DB 23; Length 23;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSVVIVGRIVL 11
DB 3 GSVVIVGRIVL 13

RESULT 10

ABG31914
ID ABG31914 standard; Peptide; 23 AA.

AC ABG31914;

DT 05-NOV-2002 (first entry)

DE KNNS4a peptide.

HCV; hepatitis C; imidazolidinone; serine protease inhibitor;
virucide; hepatotropic; antiinflammatory; NS3 protease; KNNS4a;
growth inhibitor; viral infection; blood plasma processing.

OS Synthetic.

PN WO200248157-A2.

PD 20-JUN-2002.

PF 12-DEC-2001; 2001WO-US47916.

PR 13-DEC-2000; 2000US-255168P.

PA (BRIM) BRISTOL MYERS SQUIBB PHARMA CO.

PI Han Q;

DR WPI; 2002-599498/64.

New imidazolidinones useful as serine protease inhibitors in the treatment of e.g. viral infection

Example 20; Page 112; 173pp; English.

This invention relates to novel imidazolidinones or their stereoisomers, salts or prodrugs which are useful as serine protease inhibitors. The imidazolidinones of the invention may have virucide, hepatotropic, or antiinflammatory activities and may be used as a serine protease inhibitor (preferably Hepatitis C virus (HCV) NS3 protease inhibitor) or a HCV growth inhibitor. Compounds of the invention are useful for treating viral infection e.g. hepatitis C virus (HCV) infection and as a reagent used as inhibitors of HCV protease in the processing of blood plasma for diagnostic and other commercial purposes. The imidazolidinones of the invention inhibit HCV NS3 protease and/or HCV growth and thus can be used in the blood plasma assay. The present sequence represents the KNNS4a peptide used in enzyme assay experiments in the examples of the specification.

Sequence 23 AA;

Query Match 100.0%; Score 49; DB 23; Length 23;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSVVIVGRIVL 11
DB 3 GSVVIVGRIVL 13

RESULT 11

AAE18687
ID AAE18687 standard; peptide; 23 AA.

AC AAE18687;

DT 17-MAY-2002 (first entry)

DE NS4A peptide used to purify NS3/4a conformational epitope.

Hepatitis C virus; NS3/4a antibody; HCV infection; NS4A peptide.

OS Unidentified.

PN WO200196875-A2.

PD 20-DEC-2001.

PF 14-JUN-2001; 2001WO-US19369.

PR 15-JUN-2000; 2000US-212082P.

PR 02-APR-2001; 2001US-280811P.

PR 02-APR-2001; 2001US-280867P.

PA (CHIR) CHIRON CORP.

PI Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Coit D;

PI Medina-Selby A;

DR WPI; 2002-179522/23.

Immunocassay solid support useful for detecting hepatitis C virus infection in a biological sample, comprises at least one of HCV anti-core antibody and HCV NS3/4a epitope, bound to the support

Example 2; Page 50; 87pp; English.

The present invention relates to hepatitis C virus (HCV) core antigen and NS (nonstructural) 3/4a antibody combination assay that can detect both HCV antigens and antibodies present in a sample using a single solid matrix as well as immunocassay solid supports for use in the assay. The solid support is useful for detecting HCV infection in a biological sample. The present sequence is NS4A peptide which is used to purify NS3/4a conformational epitope in the exemplification of the invention.

Sequence 23 AA;

Query Match 100.0%; Score 49; DB 23; Length 23;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSVVIVGRIVL 11
DB 3 GSVVIVGRIVL 13

RESULT 12

AAU76376
ID AAU76376 standard; Peptide; 23 AA.

AC AAU76376;

DT 08-MAY-2002 (first entry)


```

XX Hepatitis C virus (non-structural protein) NS4A peptide sequence.
DE
XX Hepatitis C virus: HCV; NS3/4a conformational epitope; seroconversion;
KW immunoassay solid support; multiple epitope fusion antigen; MEFA;
KW non-structural protein; NS4A.
XX
OS Hepatitis C virus.
XX
PN WO200196870-A2.
XX
PD 20-DEC-2001.
XX
XX 14-JUN-2001; 2001WO-US19156.
XX
XX 15-JUN-2000; 2000US-212082P.
PR 02-APR-2001; 2001US-280811P.
PR 02-APR-2001; 2001US-280867P.
XX
XX (CHIR ) CHIRON CORP.
XX
XX Chien DY, Arcangel P, Tandeske L, George-nascimento C, Coit D;
PI Medina-selby A;
PI
DR WPI; 2002-090228/12.
XX
XX Immunoassay solid support, useful for detecting hepatitis C virus
PT infection in biological sample, comprises HCV NS3/4a conformational
PT epitope and multiple epitope fusion antigen bound to the support -
XX
XX Example 3; Page 48; 92pp: English.
XX
XX The present invention relates to a new immunoassay solid support
CC consisting essentially of at least one hepatitis C virus (HCV) NS3/4a
CC conformational epitope and a multiple epitope fusion antigen (MEFA),
CC bound to the support. The NS3/4a conformational epitope and/or
CC MEFA reacts specifically with anti-HCV antibodies present in a biological
CC sample from an HCV-infected individual. The immunoassay of the invention
CC is useful for detecting hepatitis C virus infection in a biological
CC sample. The method of the invention provides a sensitive, accurate
CC diagnostic and prognostic tool to provide adequate patient care and to
CC prevent transmission of HCV by blood and by blood products, or by
CC personal contact. Use of NS3/4a conformational epitope in combination
CC with MEFA, provides a sensitive and reliable method for detecting early
CC HCV seroconversion. Use of MEFA has the added advantages of decreasing
CC masking problems, improving sensitivity in detecting antibodies by
CC allowing a greater number of epitopes on a unit surface area of
CC substrate, and improving substrate. Detection accuracy is increased and
CC the incidence of false results is reduced because of the identification
CC and the use of highly immunogenic HCV antigens which are present during
CC the early stages of HCV seroconversion. The present amino acid sequence
CC represents the non-structural protein NS4A peptide sequence. The peptide
CC was used in the invention for the purification of NS3/4a conformational
CC epitope.
XX
XX Sequence 23 AA;
SQ
Query Match 100.0%; Score 49; DB 23; Length 23;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GSVIVIGRVL 11
DB 3 GSVIVIGRVL 13
RESULT 13
ABG72264
ID ABG72264 standard; peptide; 23 AA.
XX
XX AC ABG72264;
XX
XX DT- 06-MAR-2003 (first entry)

```

```

XX Hepatitis C Virus type-1 (HCV-1) NS4a peptide.
DE
XX Immunoassay solid support; Hepatitis C Virus type-1; HCV-1;
KW NS3/4a conformational epitope; multiple epitope fusion antigen;
KW MEFA; anti-HCV antibody; NS3/4a conformational antigen;
KW HCV infection; E2 hypervariable region.
XX
XX Hepatitis C virus type 1.
XX
XX US2002146685-A1.
XX
XX 10-OCT-2002.
XX
XX 14-JUN-2001; 2001US-0881654.
XX
XX 15-JUN-2000; 2000US-212082P.
PR 02-APR-2001; 2001US-280811P.
PR 02-APR-2001; 2001US-280867P.
XX
XX (CHIE/) CHIEN D Y.
XX
XX (ARCA/) ARCANGEL P.
XX
XX (TAND/) TANDESKE L.
XX
XX (GEOR/) GEORGE-NASCIMENTO C.
XX
XX (COIT/) COIT D.
XX
XX (MEDI/) MEDINA-SELBY A.
XX
XX Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Coit D;
PI Medina-Selby A;
PI
DR WPI; 2003-147573/14.
XX
XX Immunoassay solid support for detecting Hepatitis C Virus infection in
PT biological samples, comprises Hepatitis C Virus conformational epitope
PT and multiple epitope fusion antigen -
XX
XX Example 3; Page 17; 45pp: English.
XX
XX The present invention relates to immunoassays comprising Hepatitis C
CC virus (HCV) NS3/4a conformational epitope and multiple epitope fusion
CC antigen (MEFA), bound to a solid support. The NS3/4a epitope and/or
CC the multiple epitope fusion antigen react with anti-HCV antibodies
CC present in a biological sample from an HCV-infected individual. The
CC immunoassays and methods of the invention are useful for detecting
CC HCV infection in a biological sample. The inventive immunoassay solid
CC support provides a sensitive and reliable method for detecting early
CC HCV seroconversion. The assays can detect HCV infection caused by any
CC six known genotypes of HCV. The use of the multiple epitope fusion
CC proteins decreases masking problems, improves sensitivity in detecting
CC antibodies by allowing a greater number of epitopes on a unit area
CC of substrate, and improves selectivity. The present sequence
CC representing HCV type 1 (HCV-1) NS4a peptide is used in a protease
CC enzyme activity assay in the examples of the present invention.
XX
XX Sequence 23 AA;
SQ
Query Match 100.0%; Score 49; DB 24; Length 23;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GSVIVIGRVL 11
DB 3 GSVIVIGRVL 13
RESULT 14
AAB15212
ID AAB15212 standard; protein; 195 AA.
XX
XX AC AAB15212;
XX
XX DT 19-DEC-2000 (first entry)
XX

```

DE Hepatitis C virus NS4A-NS3 fusion protease #1.
 XX Hepatitis; NS3 protease; viral replication; chronic liver disease;
 KW liver failure; liver cancer.
 XX
 OS Hepatitis C virus.
 OS Synthetic.

PN WO200040707-A1.

XX 13-JUL-2000.

XX 06-JAN-2000; 2000WO-US00345.

XX 08-JAN-1999; 99US-0115271.

PA (BRIM) BRISTOL-MYERS SQUIBB CO.

PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;

DR WPI; 2000-465976/40.

DR N-PSDB; AAA73328.

XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C

PS Example 2; Fig 10; 66pp; English.

XX The present sequence is a fusion protein created using the Hepatitis C
 CC virus (HCV) NS3 and NS4A protease enzymes. These proteins are both
 CC essential for the replication of the virus, acting to cleave its
 CC replicative proteins from the polyprotein produced from the HCV genome.
 CC Inhibitors of the two proteins should be effective as antiviral
 CC treatments of HCV infection. This is useful as HCV can lead to chronic
 CC liver disease such as cirrhosis, liver failure and liver cancer. The
 CC present invention concerns a number of NS3 mutants and NS3-NS4A fusion
 CC proteins which can be used to identify inhibitors of this type, as well
 CC as enabling structural studies of the protease and protease:inhibitor
 CC complexes.

XX Sequence 195 AA;

Query Match 100.0%; Score 49; DB 21; Length 195;

Best Local Similarity 100.0%; Pred. No. 0.56;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSVVIVGRIVL 11

DB 5 GSVVIVGRIVL 15

RESULT 15

AAB15220

ID AAB15220 standard; protein; 195 AA.

XX AAB15220;

XX 19-DEC-2000 (first entry)

DE Hepatitis C virus NS4A-NS3 fusion protease #2.

XX Hepatitis; NS3 protease; viral replication; chronic liver disease;

KW liver failure; liver cancer; mutant; mutein.

XX Hepatitis C virus.

OS Synthetic.

XX WO200040707-A1.

XX 13-JUL-2000.

XX

PF 06-JAN-2000; 2000WO-US00345.

XX

PR 08-JAN-1999; 99US-0115271.

XX

PA (BRIM) BRISTOL-MYERS SQUIBB CO.

XX

PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;

XX

DR WPI; 2000-465976/40.

DR

DR N-PSDB; AAA73329.

XX

PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C

PS Claim 23; Fig 12; 66pp; English.

XX The present sequence is a mutated version of a fusion protein created
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
 CC proteins are both essential for the replication of the virus, acting to
 CC cleave its replicative proteins from the polyprotein produced from the
 CC HCV genome. Inhibitors of the two proteins should be effective as
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A
 CC fusion proteins which can be used to identify inhibitors of this type, as
 CC well as enabling structural studies of the protease and
 CC protease:inhibitor complexes. This sequence contains the alpha-helix0-1
 CC variant.

XX Sequence 195 AA;

Query Match 100.0%; Score 49; DB 21; Length 195;

Best Local Similarity 100.0%; Pred. No. 0.56;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSVVIVGRIVL 11

DB 5 GSVVIVGRIVL 15

Search completed: August 30, 2003, 19:12:26

Job time : 3.49162 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 19:02:22 ; Search time 0.905317 Seconds
(without alignments)
1168.492 Million cell updates/sec

Title: US-09-965-594-26

Perfect score: 49

Sequence: 1 GSVVIVGRIVL 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_76.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	98.0	3010	1 GNMVTC	genome polyprotein
2	48	98.0	3010	1 GNMVTC	genome polyprotein
3	48	98.0	3010	1 A45573	genome polyprotein
4	48	98.0	3010	1 S18030	genome polyprotein
5	48	98.0	3010	1 GNMVTC	genome polyprotein
6	44	89.8	492	2 PS0126	polyprotein - hepa
7	44	89.8	876	2 PC2219	polypeptide - hepa
8	44	89.8	3011	1 GNMVTC	genome polyprotein
9	44	89.8	3011	1 S40770	genome polyprotein
10	44	89.8	3014	1 JC5620	genome polyprotein
11	43	87.8	3011	1 GNMVTC	genome polyprotein
12	42	85.7	716	2 J01366	polyprotein - hepa
13	37	75.5	718	2 G70978	probable copper-tr
14	35	71.4	574	2 T41395	probable dna polym
15	35	71.4	876	2 E89949	valine-tRNA ligase
16	34	69.4	44	2 PS0117	H-2 class I histoc
17	34	69.4	196	2 S54580	probable membrane
18	34	69.4	224	2 C46357	env polyprotein -
19	34	69.4	570	2 E71234	hypothetical prote
20	34	69.4	649	2 T05630	hypothetical prote
21	34	69.4	859	1 VCLJEW	env polyprotein pr
22	34	69.4	859	1 VCLJ22	env polyprotein pr
23	34	69.4	859	1 VCLJEW	env polyprotein pr
24	34	69.4	859	1 VCLJEW	env polyprotein pr
25	34	69.4	859	1 VCLJ22	env polyprotein pr
26	34	69.4	859	1 VCLJ22	env polyprotein pr
27	34	69.4	859	1 VCLJ22	env polyprotein pr
28	34	69.4	860	1 VCLJ24	env polyprotein pr
29	33	67.3	37	2 PS0130	H-2 class I histoc

30 33 67.3 37 2 PS0127
31 33 67.3 38 2 PS0118
32 33 67.3 39 2 A32934
33 33 67.3 142 2 PC1307
34 33 67.3 209 2 PC1306
35 33 67.3 220 2 D75611
36 33 67.3 226 2 D90908
37 33 67.3 248 2 AG3213
38 33 67.3 271 2 F95091
39 33 67.3 291 2 B97959
40 33 67.3 300 2 C85631
41 33 67.3 316 2 F69491
42 33 67.3 331 2 F89771
43 33 67.3 386 2 B71407
44 33 67.3 395 2 E90047
45 33 67.3 405 2 C83204

H-2 class I histoc
H-2 class I histoc
H-2 class I-like h
genome polyprotein
genome polyprotein
conserved hypoteth
probable tail asse
3-oxoacyl-(acyl-ca
conserved domain p
conserved hypoteth
hypothetical prote
methenyltetrahydro
lipoprotein (impor
hypothetical prote
hypothetical prote
argininosuccinate

ALIGNMENTS

RESULT 1

GNMVT

genome polyprotein - hepatitis C virus
N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome isolated from hu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 19-Jan-2001
C:Accession: A38465
R:Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, J. Virol. 65, 1105-1113, 1991
A:Title: Structure and organization of the hepatitis C virus genome isolated from hu
A:Reference number: A38465; MUID:91140698; PMID:1847440
A:Accession: A38465
A:Molecule type: genomic RNA
A:Residues: 1-3010 <TAK>
A:Cross-references: EMBL:M58335; NID:g329770; PID:AAA72945.1; PID:g329771
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruc
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MES>
F:330-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome isolated from hu
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,207

Query Match 98.0%; Score 48; DB 1; Length 3010;

Best Local Similarity 90.9%; Pred. No. 2.4;

Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GSVVIVGRIVL 11

Db 1678 GSVVIVGRIVL 1688

RESULT 2

GNMVT

genome polyprotein - hepatitis C virus (strain J)
N:Contains: capsid protein C; envelope protein M; major envelope protein E; nonstruc
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 19-Jan-2001
C:Accession: A39253; PS0086
R:Kato, N.; Hijikata, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; S
Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990
A:Title: Molecular cloning of the human hepatitis C virus genome from Japanese patie

```
A:Reference number: A39253; MUID:91088550; PMID:2175903
A:Accession: A39253
A:Molecule type: genomic RNA
A:Residues: 1-3010 <KAT>
A:Cross-references: GB:D90208; NID:g221610; PIDN:BAA14233.1; PID:g221611
R:Kato, N.; Ohkoshi, S.; Shimotohno, K.
Proc. Jpn. Acad. 65A, 219-223, 1989
A:Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence variations
A:Reference number: PS0085
A:Accession: PS0086
A:Molecule type: genomic RNA
A:Residues: 2650-2707 <KA2>
A:Experimental source: Japanese isolate
C:Comment: The cleavage sites of this polyprotein have not been determined.
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serine
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polyprotein #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,250,305,325,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2241,2242,2243,2244,2245,2246,2247,2248,2249,2250,2251,2252,2253,2254,2255,2256,2257,2258,2259,2260,2261,2262,2263,2264,2265,2266,2267,2268,2269,2270,2271,2272,2273,2274,2275,2276,2277,2278,2279,2280,2281,2282,2283,2284,2285,2286,2287,2288,2289,2290,2291,2292,2293,2294,2295,2296,2297,2298,2299,2300,2301,2302,2303,2304,2305,2306,2307,2308,2309,2310,2311,2312,2313,2314,2315,2316,2317,2318,2319,2320,2321,2322,2323,2324,2325,2326,2327,2328,2329,2330,2331,2332,2333,2334,2335,2336,2337,2338,2339,2340,2341,2342,2343,2344,2345,2346,2347,2348,2349,2350,2351,2352,2353,2354,2355,2356,2357,2358,2359,2360,2361,2362,2363,2364,2365,2366,2367,2368,2369,2370,2371,2372,2373,2374,2375,2376,2377,2378,2379,2380,2381,2382,2383,2384,2385,2386,2387,2388,2389,2390,2391,2392,2393,2394,2395,2396,2397,2398,2399,2400,2401,2402,2403,2404,2405,2406,2407,2408,2409,2410,2411,2412,2413,2414,2415,2416,2417,2418,2419,2420,2421,2422,2423,2424,2425,2426,2427,2428,2429,2430,2431,2432,2433,2434,2435,2436,2437,2438,2439,2440,2441,2442,2443,2444,2445,2446,2447,2448,2449,2450,2451,2452,2453,2454,2455,2456,2457,2458,2459,2460,2461,2462,2463,2464,2465,2466,2467,2468,2469,2470,2471,2472,2473,2474,2475,2476,2477,2478,2479,2480,2481,2482,2483,2484,2485,2486,2487,2488,2489,2490,2491,2492,2493,2494,2495,2496,2497,2498,2499,2500,2501,2502,2503,2504,2505,2506,2507,2508,2509,2510,2511,2512,2513,2514,2515,2516,2517,2518,2519,2520,2521,2522,2523,2524,2525,2526,2527,2528,2529,2530,2531,2532,2533,2534,2535,2536,2537,2538,2539,2540,2541,2542,2543,2544,2545,2546,2547,2548,2549,2550,2551,2552,2553,2554,2555,2556,2557,2558,2559,2560,2561,2562,2563,2564,2565,2566,2567,2568,2569,2570,2571,2572,2573,2574,2575,2576,2577,2578,2579,2580,2581,2582,2583,2584,2585,2586,2587,2588,2589,2590,2591,2592,2593,2594,2595,2596,2597,2598,2599,2600,2601,2602,2603,2604,2605,2606,2607,2608,2609,2610,2611,2612,2613,2614,2615,2616,2617,2618,2619,2620,2621,2622,2623,2624,2625,2626,2627,2628,2629,2630,2631,2632,2633,2634,2635,2636,2637,2638,2639,2640,2641,2642,2643,2644,2645,2646,2647,2648,2649,2650,2651,2652,2653,2654,2655,2656,2657,2658,2659,2660,2661,2662,2663,2664,2665,2666,2667,2668,2669,2670,2671,2672,2673,2674,2675,2676,2677,2678,2679,2680,2681,2682,2683,2684,2685,2686,2687,2688,2689,2690,2691,2692,2693,2694,2695,2696,2697,2698,2699,2700,2701,2702,2703,2704,2705,2706,2707,2708,2709,2710,2711,2712,2713,2714,2715,2716,2717,2718,2719,2720,2721,2722,2723,2724,2725,2726,2727,2728,2729,2730,2731,2732,2733,2734,2735,2736,2737,2738,2739,2740,2741,2742,2743,2744,2745,2746,2747,2748,2749,2750,2751,2752,2753,2754,2755,2756,2757,2758,2759,2760,2761,2762,2763,2764,2765,2766,2767,2768,2769,2770,2771,2772,2773,2774,2775,2776,2777,2778,2779,2780,2781,2782,2783,2784,2785,2786,2787,2788,2789,2790,2791,2792,2793,2794,2795,2796,2797,2798,2799,2800,2801,2802,2803,2804,2805,2806,2807,2808,2809,2810,2811,2812,2813,2814,2815,2816,2817,2818,2819,2820,2821,2822,2823,2824,2825,2826,2827,2828,2829,2830,2831,2832,2833,2834,2835,2836,2837,2838,2839,2840,2841,2842,2843,2844,2845,2846,2847,2848,2849,2850,2851,2852,2853,2854,2855,2856,2857,2858,2859,2860,2861,2862,2863,2864,2865,2866,2867,2868,2869,2870,2871,2872,2873,2874,2875,2876,2877,2878,2879,2880,2881,2882,2883,2884,2885,2886,2887,2888,2889,2890,2891,2892,2893,2894,2895,2896,2897,2898,2899,2900,2901,2902,2903,2904,2905,2906,2907,2908,2909,2910,2911,2912,2913,2914,2915,2916,2917,2918,2919,2920,2921,2922,2923,2924,2925,2926,2927,2928,2929,2930,2931,2932,2933,2934,2935,2936,2937,2938,2939,2940,2941,2942,2943,2944,2945,2946,2947,2948,2949,2950,2951,2952,2953,2954,2955,2956,2957,2958,2959,2960,2961,2962,2963,2964,2965,2966,2967,2968,2969,2970,2971,2972,2973,2974,2975,2976,2977,2978,2979,2980,2981,2982,2983,2984,2985,2986,2987,2988,2989,2990,2991,2992,2993,2994,2995,2996,2997,2998,2999,3000,3001,3002,3003,3004,3005,3006,3007,3008,3009,3010,3011,3012,3013,3014,3015,3016,3017,3018,3019,3020,3021,3022,3023,3024,3025,3026,3027,3028,3029,3030,3031,3032,3033,3034,3035,3036,3037,3038,3039,3040,3041,3042,3043,3044,3045,3046,3047,3048,3049,3050,3051,3052,3053,3054,3055,3056,3057,3058,3059,3060,3061,3062,3063,3064,3065,3066,3067,3068,3069,3070,3071,3072,3073,3074,3075,3076,3077,3078,3079,3080,3081,3082,3083,3084,3085,3086,3087,3088,3089,3090,3091,3092,3093,3094,3095,3096,3097,3098,3099,3100,3101,3102,3103,3104,3105,3106,3107,3108,3109,3110,3111,3112,3113,3114,3115,3116,3117,3118,3119,3120,3121,3122,3123,3124,3125,3126,3127,3128,3129,3130,3131,3132,3133,3134,3135,3136,3137,3138,3139,3140,3141,3142,3143,3144,3145,3146,3147,3148,3149,3150,3151,3152,3153,3154,3155,3156,3157,3158,3159,3160,3161,3162,3163,3164,3165,3166,3167,3168,3169,3170,3171,3172,3173,3174,3175,3176,3177,3178,3179,3180,3181,3182,3183,3184,3185,3186,3187,3188,3189,3190,3191,3192,3193,3194,3195,3196,3197,3198,3199,3200,3201,3202,3203,3204,3205,3206,3207,3208,3209,3210,3211,3212,3213,3214,3215,3216,3217,3218,3219,3220,3221,3222,3223,3224,3225,3226,3227,3228,3229,3230,3231,3232,3233,3234,3235,3236,3237,3238,3239,3240,3241,3242,3243,3244,3245,3246,3247,3248,3249,3250,3251,3252,3253,3254,3255,3256,3257,3258,3259,3260,3261,3262,3263,3264,3265,3266,3267,3268,3269,3270,3271,3272,3273,3274,3275,3276,3277,3278,3279,3280,3281,3282,3283,3284,3285,3286,3287,3288,3289,3290,3291,3292,3293,3294,3295,3296,3297,3298,3299,3300,3301,3302,3303,3304,3305,3306,3307,3308,3309,3310,3311,3312,3313,3314,3315,3316,3317,3318,3319,3320,3321,3322,3323,3324,3325,3326,3327,3328,3329,3330,3331,3332,3333,3334,3335,3336,3337,3338,3339,3340,3341,3342,3343,3344,3345,3346,3347,3348,3349,3350,3351,3352,3353,3354,3355,3356,3357,3358,3359,3360,3361,3362,3363,3364,3365,3366,3367,3368,3369,3370,3371,3372,3373,3374,3375,3376,3377,3378,3379,3380,3381,3382,3383,3384,3385,3386,3387,3388,3389,3390,3391,3392,3393,3394,3395,3396,3397,3398,3399,3400,3401,3402,3403,3404,3405,3406,3407,3408,3409,3410,3411,3412,3413,3414,3415,3416,3417,3418,3419,3420,3421,3422,3423,3424,3425,3426,3427,3428,3429,3430,3431,3432,3433,3434,3435,3436,3437,3438,3439,3440,3441,3442,3443,3444,3445,3446,3447,3448,3449,3450,3451,3452,3453,3454,3455,3456,3457,3458,3459,3460,3461,3462,3463,3464,3465,3466,3467,3468,3469,3470,3471,3472,3473,3474,3475,3476,3477,3478,3479,3480,3481,3482,3483,3484,3485,3486,3487,3488,3489,3490,3491,3492,3493,3494,3495,3496,3497,3498,3499,3500,3501,3502,3503,3504,3505,3506,3507,3508,3509,3510,3511,3512,3513,3514,3515,3516,3517,3518,3519,3520,3521,3522,3523,3524,3525,3526,3527,3528,3529,3530,3531,3532,3533,3534,3535,3536,3537,3538,3539,3540,3541,3542,3543,3544,3545,3546,3547,3548,3549,3550,3551,3552,3553,3554,3555,3556,3557,3558,3559,3560,3561,3562,3563,3564,3565,3566,3567,3568,3569,3570,3571,3572,3573,3574,3575,3576,3577,3578,3579,3580,3581,3582,3583,3584,3585,3586,3587,3588,3589,3590,3591,3592,3593,3594,3595,3596,3597,3598,3599,3600,3601,3602,3603,3604,3605,3606,3607,3608,3609,3610,3611,3612,3613,3614,3615,3616,3617,3618,3619,3620,3621,3622,3623,3624,3625,3626,3627,3628,3629,3630,3631,3632,3633,3634,3635,3636,3637,3638,3639,3640,3641,3642,3643,3644,3645,3646,3647,3648,3649,3650,3651,3652,3653,3654,3655,3656,3657,3658,3659,3660,3661,3662,3663,3664,3665,3666,3667,3668,3669,3670,3671,3672,3673,3674,3675,3676,3677,3678,3679,3680,3681,3682,3683,3684,3685,3686,3687,3688,3689,3690,3691,3692,3693,3694,3695,3696,3697,3698,3699,3700,3701,3702,3703,3704,3705,3706,3707,3708,3709,3710,3711,3712,3713,3714,3715,3716,3717,3718,3719,3720,3721,3722,3723,3724,3725,3726,3727,3728,3729,3730,3731,3732,3733,3734,3735,3736,3737,3738,3739,3740,3741,3742,3743,3744,3745,3746,3747,3748,3749,3750,3751,3752,3753,3754,3755,3756,3757,3758,3759,3760,3761,3762,3763,3764,3765,3766,3767,3768,3769,3770,3771,3772,3773,3774,3775,3776,3777,3778,3779,3780,3781,3782,3783,3784,3785,3786,3787,3788,3789,3790,3791,3792,3793,3794,3795,3796,3797,3798,3799,3800,3801,3802,3803,3804,3805,3806,3807,3808,3809,3810,3811,3812,3813,3814,3815,3816,3817,3818,3819,3820,3821,3822,3823,3824,3825,3826,3827,3828,3829,3830,3831,3832,3833,3834,3835,3836,3837,3838,3839,3840,3841,3842,3843,3844,3845,3846,3847,3848,3849,3850,3851,3852,3853,3854,3855,3856,3857,3858,3859,3860,3861,3862,3863,3864,3865,3866,3867,3868,3869,3870,3871,3872,3873,3874,3875,3876,3877,3878,3879,3880,3881,3882,3883,3884,3885,3886,3887,3888,3889,3890,3891,3892,3893,3894,3895,3896,3897,3898,3899,3900,3901,3902,3903,3904,3905,3906,3907,3908,3909,3910,3911,3912,3913,3914,3915,3916,3917,3918,3919,3920,3921,3922,3923,3924,3925,3926,3927,3928,3929,3930,3931,3932,3933,3934,3935,3936,3937,3938,3939,3940,3941,3942,3943,3944,3945,3946,3947,3948,3949,3950,3951,3952,3953,3954,3955,3956,3957,3958,3959,3960,3961,3962,3963,3964,3965,3966,3967,3968,3969,3970,3971,3972,3973,3974,3975,3976,3977,3978,3979,3980,3981,3982,3983,3984,3985,3986,3987,3988,3989,3990,3991,3992,3993,3994,3995,3996,3997,3998,3999,4000,4001,4002,4003,4004,4005,4006,4007,4008,4009,4010,4011,4012,4013,4014,4015,4016,4017,4018,4019,4020,4021,4022,4023,4024,4025,4026,4027,4028,4029,4030,4031,4032,4033,4034,4035,4036,4037,4038,4039,4040,4041,4042,4043,4044,4045,4046,4047,4048,4049,4050,4051,4052,4053,4054,4055,4056,4057,4058,4059,4060,4061,4062,4063,4064,4065,4066,4067,4068,4069,4070,4071,4072,4073,4074,4075,4076,4077,4078,4079,4080,4081,4082,4083,4084,4085,4086,4087,4088,4089,4090,4091,4092,4093,4094,4095,4096,4097,4098,4099,4100,4101,4102,4103,4104,4105,4106,4107,4108,4109,4110,4111,4112,4113,4114,4115,4116,4117,4118,4119,4120,4121,4122,4123,4124,4125,4126,4127,4128,4129,4130,4131,4132,4133,4134,4135,4136,4137,4138,4139,4140,4141,4142,4143,4144,4145,4146,4147,4148,4149,4150,4151,4152,4153,4154,4155,4156,4157,4158,4159,4160,4161,4162,4163,4164,4165,4166,4167,4168,4169,4170,4171,4172,4173,4174,4175,4176,4177,4178,4179,4180,4181,4182,4183,4184,4185,4186,4187,4188,4189,4190,4191,4192,4193,4194,4195,4196,4197,4198,4199,4200,4201,4202,4203,4204,4205,4206,4207,4208,4209,4210,4211,4212,4213,4214,4215,4216,4217,4218,4219,4220,4221,4222,4223,4224,4225,4226,4227,4228,4229,4230,4231,4232,4233,4234,4235,4236,4237,4238,4239,4240,4241,4242,4243,4244,4245,4246,4247,4248,4249,4250,4251,4252,4253,4254,4255,4256,4257,4258,4259,4260,4261,4262,4263,4264,4265,4266,4267,4268,4269,4270,4271,4272,4273,4274,4275,4276,4277,4278,4279,4280,4281,4282,4283,4284,4285,4286,4287,4288,4289,4290,4291,4292,4293,4294,4295,4296,4297,4298,4299,4300,4301,4302,4303,4304,4305,4306,4307,4308,4309,4310,4311,4312,4313,4314,4315,4316,4317,4318,4319,4320,4321,4322,4323,4324,4325,4326,4327,4328,4329,4330,4331,4332,4333,4334,4335,4336,4337,4338,4339,4340,4341,4342,4343,4344,4345,4346,4347,4348,4349,4350,4351,4352,4353,4354,4355,4356,4357,4358,4359,4360,4361,4362,4363,4364,4365,4366,4367,4368,4369,4370,4371,4372,4373,4374,4375,4376,4377,4378,4379,4380,4381,4382,4383,4384,4385,4386,4387,4388,4389,4390,4391,4392,4393,4394,4395,4396,4397,4398,4399,4400,4401,4402,4403,4404,4405,4406,4407,4408,4409,4410,4411,4412,4413,4414,4415,4416,4417,4418,4419,4420,4421,4422,4423,4424,4425,4426,4427,4428,4429,4430,4431,4432,4433,4434,4435,4436,4437,4438,4439,4440,4441,4442,4443,4444,4445,4446,4447,4448,4449,4450,4451,4452,4453,4454,4455,4456,4457,4458,4459,4460,4461,4462,4463,4464,4465,4466,4467,4468,4469,4470,4471,4472,4473,4474,4475,4476,4477,4478,4479,4480,4481,4482,4483,4484,4485,4486,4487,4488,4489,4490,4491,4492,4493,4494,4495,4496,4497,4498,4499,4500,4501,4502,4503,4504,4505,4506,4507,4508,4509,4510,4511,4512,4513,4514,4515,4516,4517,4518,4519,4520,4521,4522,4523,4524,4525,4526,4527,4528,4529,4530,4531,4532,4533,4534,4535,4536,4537,4538,4539,4540,4541,4542,4543,4544,4545
```

R:Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.
 Virology 188, 102-113, 1992
 A:Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the
 A:Reference number: A40244; MUID:92230206; PMID:1314449
 A:Accession: A40244
 A:Molecule type: genomic RNA
 A:Residues: 1-3010 <CHE>
 A:Cross-references: GB:M84754
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
 F:1-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: major envelope protein E #status predicted <EPW>
 F:192-389/Product: major envelope protein E #status predicted <NEE>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: hepatitis C virus genome polyprotein NS2 #status predicted <NS2>
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B
 F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,233,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,207

Query Match 98.0%; Score 48; DB 1; Length 3010;
 Best Local Similarity 90.9%; Pred. No. 2.4;
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GSVVIVGRIVL 11
 |||||
 Db 1678 GSVVIVGRIVL 1688

RESULT 6
 PS0326
 polyprotein - hepatitis C virus (isolate Fla) (fragments)
 C:Species: hepatitis C virus
 C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
 C:Accession: PS0326
 R:Li, J.S.; Tong, S.P.; Vitvitski, L.; Lepot, D.; Trepo, C.
 Gene 105, 167-172, 1991
 A:Title: Two French genotypes of hepatitis C virus: homology of the predominant genotype
 A:Reference number: PS0326; MUID:92039028; PMID:1718820
 A:Accession: PS0326
 A:Molecule type: genomic RNA
 A:Residues: 1-492 <LIJ>
 A:Cross-references: GB:M60220
 A:Note: this sequence corresponds to nonstructural protein NS3 region
 A:Note: translation of the nucleotide sequence is not complete
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: polyprotein

Query Match 89.8%; Score 44; DB 2; Length 492;
 Best Local Similarity 90.9%; Pred. No. 2.3;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GSVVIVGRIVL 11
 |||||
 Db 229 GSVVIVGRIVL 239

RESULT 7
 PC2219
 polyprotein - hepatitis C virus (type 5a) (fragments)
 N:Contains: core protein; E1 (carboxyl end); E2/NS1 (amino end); NS3 protein; NS4A prote
 C:Species: hepatitis C virus
 C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 17-Nov-2000
 C:Accession: PC2219
 R:Stuyver, L.; Arnhen, W.V.; Wyseur, A.; Maertens, G.
 Biochem. Biophys. Res. Commun. 202, 1308-1314, 1994
 A:Title: Cloning and phylogenetic analysis of the core, E2, and NS3/NS4 regions of the h
 A:Reference number: PC2219; MUID:94338342; PMID:7520237
 A:Accession: PC2219

A:Molecule type: mRNA
 A:Residues: 1-876 <STU>
 A:Cross-references: GB:L29577; GB:L29578; GB:L29579
 A:Experimental source: serum
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: glycoprotein
 F:1-191/Product: core #status predicted <COE>
 F:68-78/Region: variable
 F:192-247/Product: E1 (carboxyl end) #status predicted <ERE>
 F:248-411/Product: E2/NS1 (amino end) #status predicted <ENR>
 F:248-338/Region: E2
 F:339-411/Region: NS1 (amino end)
 F:412-783/Product: NS3 #status predicted <NSR>
 F:784-837/Product: NS4A #status predicted <NSA>
 F:838-876/Product: NS4B #status predicted <NSB>
 F:281,287,294,312,340/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 89.8%; Score 44; DB 2; Length 876;
 Best Local Similarity 81.8%; Pred. No. 4;
 Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GSVVIVGRIVL 11
 |||||
 Db 804 GSVVIVGRIVL 814

RESULT 8
 GNVVCH
 genome polyprotein - hepatitis C virus (strain H)
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain H) (nonstr
 Protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 A:Note: host Homo sapiens (man)
 C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
 C:Accession: A36814; A41546
 R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
 submitted to Genbank, July 1992
 A:Description: Genomic structure of the human prototype strain H of hepatitis C viru
 A:Reference number: A36814
 A:Accession: A36814
 A:Molecule type: genomic RNA
 A:Residues: 1-3011 <INC>
 A:Cross-references: GB:M67463; NID:9329737; PIDN:AAA45534.1; PID:G329738
 R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
 Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
 A:Title: Genomic structure of the human prototype strain H of hepatitis C virus: com
 A:Reference number: A41546; MUID:92052256; PMID:1658800
 A:Contents: annotation
 A:Note: neither amino acid nor nucleotide sequence is given
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruc
 F:1-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: major envelope protein E #status predicted <EPW>
 F:192-389/Product: major envelope protein E #status predicted <NEE>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: hepatitis C virus genome polyprotein NS2 #status predicted <NS2>
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B
 F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
 F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,224

Query Match 89.8%; Score 44; DB 1; Length 3011;
 Best Local Similarity 90.9%; Pred. No. 13;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GSVVIVGRIVL 11
 |||||
 Db 1678 GSVVIVGRIVL 1688

RESULT 9
 S4070 genome polyprotein - hepatitis C virus
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
 C:Accession: S40770; PC1285
 R:Okamoto, H.
 submitted to the EMBL Data Library, March 1992
 A:Reference number: S40770
 A:Accession: S40770
 A:Molecule type: genomic RNA
 A:Residues: 1-3011 <OKA>
 A:Cross-references: EMBL:D10749; NID:g221586; PIDN:BAA01582.1; PID:g221587
 R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda, Jpn. J. Exp. Med. 60, 167-177, 1990
 A:Title: The 5'-terminal sequence of the hepatitis C virus genome.
 A:Reference number: PC1284; MUID:91013116; PMID:2170712
 A:Accession: PC1285
 A:Molecule type: genomic RNA
 A:Residues: 1-513 <OK2>
 A:Cross-references: GB:D00831; NID:g221511; PIDN:BAA00705.1; PID:g221512
 A:Experimental source: isolate HC-J1
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
 F:116-191/Product: capsid protein C #status predicted <CPC>
 F:192-389/Product: envelope protein M #status predicted <EPM>
 F:390-729/Product: major envelope protein E #status predicted <MEE>
 F:730-1006/Product: nonstructural protein NS1 #status predicted <NS1>
 F:1007-1615/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1230-1237/Product: hepatitis C virus genome polyprotein NS5 #status predicted <NS5>
 F:1312-1317/Product: nucleotide-binding motif A (P-loop)
 F:1316-1319/Product: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
 F:1863-2013/Product: nonstructural protein NS5 #status predicted <NS5>
 F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
 F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
 Query Match 89.8%; Score 44; DB 1; Length 3011;
 Best Local Similarity 90.9%; Pred. No. 13;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 GSVVIVGRVIL 11
 | | | | | | | | | |
 Db 1678 GCVVIVGRVIL 1688

RESULT 10
 JC5620 genome polyprotein - hepatitis C virus (isolate EUH1480)
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
 C:Accession: JC5620
 R:Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.
 Biochem. Biophys. Res. Commun. 236, 44-49, 1997
 A:Title: The complete coding sequence of hepatitis C virus genotype 5a, the predominant
 A:Reference number: JC5620; MUID:97366593; PMID:9223423
 A:Accession: JC5620
 A:Molecule type: mRNA
 A:Residues: 1-3014 <CHA>
 A:Cross-references: GB:Y13184
 A:Experimental source: genotype 5a, which predominates in South Africa
 A:Note: The translation of the nucleotide sequence is not complete in this paper
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
 F:2-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <MEE>
 F:384-408/Product: hypervariable #status predicted

F:390-730/Product: nonstructural protein NS1 #status predicted <NS1>
 F:731-1007/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1008-1616/Product: hepatitis C virus NS4b; nonstructural protein NS5
 F:1231-1238/Product: nucleotide-binding motif A (P-loop)
 F:1313-1318/Product: DEXH motif
 F:1317-1320/Product: DEXH motif
 F:1617-1863/Product: nonstructural protein NS4a #status predicted <NS4a>
 F:1864-2014/Product: nonstructural protein NS5 #status predicted <NS5>
 F:2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>
 F:2210-2249/Product: interferon sensitivity determining #status predicted
 Query Match 89.8%; Score 44; DB 1; Length 3014;
 Best Local Similarity 81.8%; Pred. No. 13;
 Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 GSVVIVGRVIL 11
 | | | | | | | | | |
 Db 1679 GSAIVGRVIL 1689

RESULT 11
 GNVVC3 genome polyprotein - hepatitis C virus (strain HCV-1)
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 19-Jan-2001
 C:Accession: A39166; PQ0403; PQ0404
 R:Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991
 A:Title: Genetic organization and diversity of the hepatitis C virus.
 A:Reference number: A39166; MUID:91172826; PMID:1848704
 A:Accession: A39166
 A:Molecule type: mRNA
 A:Residues: 1-3011 <CHO>
 A:Cross-references: GB:M62321; NID:g329873; PIDN:AAA45676.1; PID:g329874
 R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, J. Gen. Virol. 73, 1131-1141, 1992
 A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship
 A:Reference number: PQ0393; MUID:92268871; PMID:1316959
 A:Accession: PQ0403
 A:Molecule type: genomic RNA
 A:Residues: 1577-1633 <CHA>
 A:Cross-references: DDBJ:D10128
 A:Experimental source: isolates E-b16
 A:Accession: PQ0404
 A:Status: preliminary
 A:Molecule type: genomic RNA
 A:Residues: 1577-1633 <CH2>
 A:Experimental source: isolates E-b17
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruct
 F:1-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <MEE>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: hepatitis C virus genome polyprotein NS5 #status predicted <NS5>
 F:1230-1237/Product: hepatitis C virus genome polyprotein NS5 #status predicted <NS5>
 F:1312-1317/Product: DEXH motif
 F:1316-1319/Product: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
 F:1863-2013/Product: nonstructural protein NS5 #status predicted <NS5>
 F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
 F:195-209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077

Query Match 87.8%; Score 43; DB 1; Length 3011;
 Best Local Similarity 81.8%; Pred. No. 20;
 Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 GSVVIVGRVIL 11
 | | | | | | | | | |
 Db 1678 GCVVIVGRVIL 1688

RESULT 12

polyprotein - hepatitis C virus (French isolate) (fragments)
 C:Species: hepatitis C virus
 C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
 C:Accession: J01366
 R:Kremendorf, D.; Porchon, C.; Kim, J.P.; Reyes, G.R.; Brechot, C.
 J. gen. Virol. 72, 2557-2561, 1991
 A:Title: Partial nucleotide sequence analysis of a French hepatitis C virus: implication
 A:Reference number: J01366; MUID:92013977; PMID:1655961

A:Accession: J01366
 A:Molecule type: genomic RNA
 A:Residues: 1-716 <KRE>
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: glycoprotein; polyprotein
 F:84,90,97,115,143,199,223,243,290,312/Binding site: carbohydrate (Asn) (covalent) #stat

Query Match 85.7%; Score 42; DB 2; Length 716;
 Best Local Similarity 72.7%; Pred. No. 7.6;
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GSVVIVGRIVL 11

DB 627 GCVVIVGRVIL 637

RESULT 13

probable copper-transporting atpase 11/9 - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 18-Aug-2000
 C:Accession: G70978
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: G70978
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA
 A:Residues: 1-718 <COL>
 A:Cross-references: GB:292771; GB:AL123456; NID:93242259; PIDN:CAB07083.1; PID:gl877325
 F:128-434/Domain: ATPase transduction domain homology <ATP>
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: ctpC
 C:Superfamily: Enterococcus copper-transporting ATPase copB; ATPase nucleotide-binding d

F:532-676/Domain: ATPase nucleotide-binding domain homology <ATN>

Query Match 75.5%; Score 37; DB 2; Length 718;
 Best Local Similarity 63.6%; Pred. No. 63;
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 GSVVIVGRIVL 11

DB 286 GSVVIVGRVV 296

RESULT 14

probable dna polymerase alpha-primase associated subunit - fission yeast (Schizosacchar

C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 C:Accession: T41395
 R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Murphy, L.; Harris, D.
 submitted to the EMBL Data Library, May 1998
 A:Reference number: 221991
 A:Accession: T41395
 A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-574 <WOO>

A:Cross-references: EMBL:AL023704; PIDN:CAAL9261.1; GSPDB:GN00068; SPDB:SPCC553.09c

A:Experimental source: strain 972h-; cosmid c553

C:Genetics:

A:Gene: SPOB:SPCC553.09c

A:Map position: 3

A:Introns: 89/2; 415/3; 518/3

Query Match 71.4%; Score 35; DB 2; Length 574;
 Best Local Similarity 77.8%; Pred. No. 1.2e+02;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 VVIVGRIVL 11

DB 205 VVVVGRIVV 213

RESULT 15

E89949

valine-tRNA ligase [imported] - Staphylococcus aureus (strain N315)

C:Species: Staphylococcus aureus

C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001

C:Accession: E89949

R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Cui, A.; Mizutani-Oi, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, T.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
 Lancet 357, 1225-1240, 2001

A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A:Reference number: A89758; MUID:21311952; PMID:11418146

A:Accession: E89949

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-876 <KUR>

A:Cross-references: GB:BA000018; PID:gl3701460; PIDN:BAB42754.1; GSPDB:GN00149

A:Experimental source: strain N315

C:Genetics:

A:Gene: valS

C:Superfamily: valine-tRNA ligase

Query Match 71.4%; Score 35; DB 2; Length 876;
 Best Local Similarity 70.0%; Pred. No. 1.8e+02;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 SVVIVGRIVL 11

DB 794 SVVIAGKVVL 803

Search completed: August 30, 2003, 19:20:34

Job time : 2.90532 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 18:01:52 ; Search time 0.544792 Seconds
(without alignments)
949.524 Million cell updates/sec

Title: US-09-965-594-26
Perfect score: 49
Sequence: 1 GSVVIVGRVL 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	98.0	3010	POLG_HCVBK	P26663 h genome po
2	48	98.0	3010	POLG_HCVJA	P26662 h genome po
3	48	98.0	3010	POLG_HCVJT	Q00369 h genome po
4	48	98.0	3010	POLG_HCVTW	P29846 h genome po
5	44	89.8	3011	POLG_HCVH	P27958 h genome po
6	43	87.8	3011	POLG_HCVI	P26664 h genome po
7	37	75.5	718	CTPC_MYCTU	P96875 mycobacteri
8	35	71.4	574	DPO2_SCHPO	Q74946 schizosacch
9	34	69.4	196	YM07_YEAST	Q04487 saccharomyc
10	34	69.4	260	MCH_METTR	Q94847 methylosinu
11	34	69.4	859	ENV_EIAV1	P22427 equine infe
12	34	69.4	859	ENV_EIAV2	P22428 equine infe
13	34	69.4	859	ENV_EIAV3	P22429 equine infe
14	34	69.4	859	ENV_EIAV9	P11306 equine infe
15	34	69.4	859	ENV_EIAVC	P32541 equine infe
16	34	69.4	859	ENV_EIAVM	P16082 equine infe
17	34	69.4	859	ENV_EIAVY	P06751 equine infe
18	34	69.4	860	ENV_EIAVY	P22430 equine infe
19	33	67.3	316	MCH_ARCFU	O28344 archaeoglob
20	33	67.3	405	ASSY_PSEAE	O94844 pseudomonas
21	33	67.3	456	SHU7_ECOLI	P09751 escherichia
22	33	67.3	457	CYSG_ECOLI	P11098 escherichia
23	33	67.3	560	J1160_HORVU	Q00531 hordeum vul
24	33	67.3	725	CTPC_MYCLE	Q90011 mycobacteri
25	33	67.3	968	MM12_MYCTU	Q11171 mycobacteri
26	33	67.3	1158	CND1_SCHPO	O94679 schizosacch
27	32	65.3	249	COBM_RHOER	O53138 rhodococcus
28	32	65.3	332	PLSX_THETN	O8r946 thermoaer
29	32	65.3	388	Y8C8_SALT	O8z455 salmonella
30	32	65.3	411	DBOB_LACLA	Q9ch12 lactococcus
31	32	65.3	411	DBOB_LACLC	O32808 lactococcus
32	32	65.3	417	HS47_HUMAN	P29043 homo sapien
33	32	65.3	417	HS47_MOUSE	P19324 mus musculu

34 32 65.3 417 1 HS47_RAT P29457 rattus norv
35 32 65.3 418 1 CBP2_HUMAN P50454 homo sapien
36 32 65.3 509 1 YFCC_HAEIN P44023 haemophilus
37 32 65.3 752 1 CTPB_MYCTU Q10877 mycobacteri
38 32 65.3 855 1 YB29_YEAST P38297 saccharomyc
39 31 63.3 84 1 RL27_CHLTE Q8kcb7 chlorobium
40 31 63.3 193 1 YNES_BACSU Q45064 bacillus su
41 31 63.3 223 1 VTAL_LAMBD P03730 bacteriophia
42 31 63.3 228 1 NEUA_HAEIN Q57140 haemophilus
43 31 63.3 231 1 RS2_SULSO P95993 sulfolobus
44 31 63.3 251 1 FGFN_RAT Q8v182 rattus norv
45 31 63.3 258 1 UPKA_BOVIN P38572 bos taurus

ALIGNMENTS

RESULT 1
POLG_HCVBK STANDARD; PRT; 3010 AA.
AC P26663;
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DI 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P36); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate BK) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.
OX NCBI_TaxID=11105;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-91140698; PubMed-1847440;
RA Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J., Onishi E., Andoh T., Yoshida I., Okayama H.;
RT "Structure and organization of the hepatitis C virus genome isolated from human carriers";
RL J. Virol. 65:1105-1113 (1991).
[2]
RP SEQUENCE OF 1487-1500.
RX MEDLINE-96235224; PubMed-8647104;
RA Borowski P., Heiland M., Oehlmann K., Becker B., Kornetky L.;
RT "Non-structural protein 3 of hepatitis C virus inhibits phosphorylation mediated by cAMP-dependent protein kinase";
RL Eur. J. Biochem. 237:611-618 (1996).
[3]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
RX MEDLINE-97015088; PubMed-8861916;
RA Love R.A., Parge H.E., Wickersham J.A., Hostomsky Z., Habuka N., Moomaw E.W., Adachi T., Hostomsky Z.;
RT "The crystal structure of hepatitis C virus NS3 proteinase reveals a trypsin-like fold and a structural zinc binding site";
RL Cell 87:331-342 (1996).
[4]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
RX MEDLINE-98227846; PubMed-9568891;
RA Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M., Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;
RT "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C virus: a 2.2-A resolution structure in a hexagonal crystal form";
RL Protein Sci. 7:837-847 (1998).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +

(RNA)(N).
 -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
 -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

 CC This Swiss-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 CC EMBL: M58335; AAA72945.1; --
 CC PIR: A38465; GNVVTC.
 CC PDB: 1A10; 25-MAR-98.
 CC PDB: 1JXP; 14-JAN-98.
 CC PDB: 1NS3; 08-APR-98.
 CC PDB: 1C2P; 15-NOV-00.
 CC PDB: 1CSJ; 08-NOV-99.
 CC PDB: 1GX5; 09-APR-02.
 CC PDB: 1GX6; 10-APR-02.
 CC PDB: 1QV6; 26-JUN-00.
 CC PDB: 80HM; 20-APR-99.
 CC MEROPS: S29.001; --
 CC MEROPS: U39.001; --
 CC InterPro: IPR001410; DEAD.
 CC InterPro: IPR002522; HCV_capsid.
 CC InterPro: IPR002521; HCV_core.
 CC InterPro: IPR002519; HCV_env.
 CC InterPro: IPR002531; HCV_NS1.
 CC InterPro: IPR002518; HCV_NS2.
 CC InterPro: IPR004109; HCV_NS3.
 CC InterPro: IPR000745; HCV_NS4a.
 CC InterPro: IPR001490; HCV_NS4b.
 CC InterPro: IPR002868; HCV_NS5a.
 CC InterPro: IPR002166; HCV_RdRP.
 CC InterPro: IPR007095; RNA_pol_DS_PS.
 CC InterPro: IPR007094; RNA_pol_PSVir.
 CC Pfam: PF01543; HCV_capsid; 1.
 CC Pfam: PF01542; HCV_core; 1.
 CC Pfam: PF01539; HCV_env; 1.
 CC Pfam: PF01560; HCV_NS1; 1.
 CC Pfam: PF01538; HCV_NS2; 1.
 CC Pfam: PF02907; HCV_NS3; 1.
 CC Pfam: PF01006; HCV_NS4a; 1.
 CC Pfam: PF01001; HCV_NS4b; 1.
 CC Pfam: PF01506; HCV_NS5a; 1.
 CC Pfam: PF00998; Viral_RdRP; 1.
 CC ProDom: PD186062; HCV_NS1; 1.
 CC SMART: SM00487; DEXDC; 1.
 CC PolyProtein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 CC Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 CC Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 CC 3D-structure.
 CC INIT_MET 1 1
 CC CHAIN 1 115
 CC CHAIN 116 191
 CC CHAIN 192 383
 CC CHAIN 384 729
 CC CHAIN 730 1006
 CC CHAIN 1007 1615
 CC CHAIN 1616 1862
 CC CHAIN 1863 2013
 CC CHAIN 2014 3010
 CC CHAIN 3010 369
 CC TRANSMEM 347 369
 CC ACT_SITE 1083 1083
 CC ACT_SITE 1107 1107
 CC ACT_SITE 1165 1165
 CC NP_BIND 1230 1237

FT SITE 1316 1319 DECH_BOX
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT STRAND 1031 1035
 FT HELIX 1039 1047
 FT STRAND 1050 1050
 FT STRAND 1059 1063
 FT STRAND 1068 1074
 FT TURN 1075 1076
 FT STRAND 1077 1081
 FT HELIX 1082 1085
 FT TURN 1086 1087
 FT STRAND 1090 1092
 FT TURN 1093 1094
 FT STRAND 1095 1097
 FT STRAND 1101 1103
 FT TURN 1104 1107
 FT STRAND 1108 1112
 FT STRAND 1120 1120
 FT STRAND 1122 1122
 FT STRAND 1129 1133
 FT TURN 1135 1136
 FT STRAND 1139 1144
 FT STRAND 1149 1157
 FT HELIX 1158 1161
 FT TURN 1162 1163
 FT TURN 1165 1166
 FT STRAND 1168 1171
 FT TURN 1172 1174
 FT STRAND 1175 1186
 FT TURN 1187 1188
 FT STRAND 1189 1197
 FT HELIX 1198 1202
 FT TURN 1203 1204
 FT STRAND 1260 1268

Query Match 98.0%; Score 48; DB 1; Length 3010;
 Best Local Similarity 90.9%; Pred. No. 1.1;
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GSVIVGRVL 11
 Db 1678 GSVIVGRIL 1688

RESULT 2
 POLG_HCVJA STANDARD; PRT: 3010 AA.
 ID POLG_HCVJA AC P26662;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2

DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)).
 OS Hepatitis C virus (isolate Japanese) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=91088550; PubMed=2175903;
 RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
 RA Sugimura T., Shimotohno K.;
 RT "Molecular cloning of the human hepatitis C virus genome from
 RT Japanese patients with non-A, non-B hepatitis.";
 EL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
 RN [2]
 RP DISCUSSION OF SEQUENCE.
 RA MEDLINE=91192160; PubMed=1849489;
 RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraliso K.,
 RA Ohkoshi S., Shimotohno K.;
 RT "Molecular structure of the Japanese hepatitis C viral genome.";
 EL FEBS Lett. 280:325-328(1991).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC [RNA](N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; D90208; BAA14233.1; -;
 DR PIR; A39253; GNWVCJ.
 DR HSSP; P26663; LUXP.
 DR MEROPS; S29.001; -;
 DR MEROPS; U39.001; -;
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR004109; HCV_NS3.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_NS5b.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.

DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00271; helicase_C; 1.
 DR Pfam; PF00998; Viral_RdRP; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT CHAIN 3011 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 532 532
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2788 2788
 SQ SEQUENCE 3010 AA; 327017 MW; AA993794F46DB185 CRC64;
 Query Match 98.0%; Score 48; DB 1; Length 3010;
 Best Local Similarity 90.9%; Pred. No. 1.1;
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GSVVIVGRIVL 11
 Db 1678 GSVVIVGRIVL 1688
 RESULT 3
 POLG_HCVJT
 ID POLG_HCVJT STANDARD; PRT; 3010 AA.
 AC Q00269;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 DE NS5A (P56); Nonstructural protein NS5A (P56); Nonstructural protein
 OS Hepatitis C virus (isolate HC-JT) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=31642;
 RN [1]

SEQUENCE FROM N.A.
 MEDLINE-92295714: PubMed-1318627:
 Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,
 Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
 "Molecular cloning of hepatitis C virus genome from a single Japanese
 carrier: sequence variation within the same individual and among
 infected individuals.";
 Virus Res. 23:39-53(1992).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC (RNA)(N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D11168; BAA01943.1; -
 CC PIR: A45573; A45573.
 CC PDB: 1A10; 25-MAR-98.
 CC PDB: 1JXP; 14-JAN-98.
 CC MEROPS: S29.001; -
 CC MEROPS: U39.001; -
 CC InterPro: IPR001410; DEAD.
 CC InterPro: IPR002522; HCV_capsid.
 CC InterPro: IPR002521; HCV_core.
 CC InterPro: IPR002519; HCV_gn.
 CC InterPro: IPR002531; HCV_NS1.
 CC InterPro: IPR002518; HCV_NS2.
 CC InterPro: IPR004109; HCV_NS3.
 CC InterPro: IPR000745; HCV_NS4a.
 CC InterPro: IPR001490; HCV_NS4b.
 CC InterPro: IPR002868; HCV_NS5a.
 CC InterPro: IPR002166; HCV_RGRP.
 CC InterPro: IPR007095; RNA_pol_DS_PS.
 CC InterPro: IPR007094; RNA_pol_Psvir.
 CC Pfam: PF01543; HCV_core; 1.
 CC Pfam: PF01542; HCV_core; 1.
 CC Pfam: PF01539; HCV_env; 1.
 CC Pfam: PF01560; HCV_NS1; 1.
 CC Pfam: PF01538; HCV_NS2; 1.
 CC Pfam: PF02907; HCV_NS3; 1.
 CC Pfam: PF01006; HCV_NS4a; 1.
 CC Pfam: PF01001; HCV_NS4b; 1.
 CC Pfam: PF01506; HCV_NS5a; 1.
 CC Pfam: PF00271; helicase_C; 1.
 CC Pfam: PF00998; Viral_RdRp; 1.
 CC ProDom: PD186062; HCV_NS1; 1.
 CC SMART: SM00487; DEXDC; 1.
 CC Polyprotein: Glycoprotein; Transferase; RNA-directed RNA polymerase;
 CC Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 CC Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 CC 3D-structure.
 CC INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 CC CELLULAR AMINOPEPTIDASE.
 CC CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).
 CC CHAIN 116 191 MATRIX ENVELOPE (POTENTIAL).
 CC CHAIN 192 393 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 CC CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 CC CHAIN 730 1006 NON-STRUCTURAL PROTEIN NS2 (POTENTIAL).

FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT TRANSMEM 347 369 POTENTIAL.
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 136 196 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 3010 AA; 326573 MW; 94A1C77435D642BB CRC64;
 Query Match 98.0%; Score 48; DB 1; Length 3010;
 Best Local Similarity 90.9%; Pred. No. 1.1;
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GSVWIVGRVL 11
 Db 1678 GSVWIVGRIL 1688
 RESULT 4
 POLG_HCVTW STANDARD; PRT: 3010 AA.
 ID AC P29846;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate Taiwan) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=31645;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92230206; PubMed-1314449;
 RA Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;
 RT "The Taiwanese hepatitis C virus genome: sequence determination and
 RT mapping the 5' terminus of viral genomic and antigenomic RNA.";
 RL Virology 188:102-113(1992).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +

CC [RNA](N).

CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

CC PROTEIN C AND RNA.

CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: M84754; .. NOT_ANNOTATED_CDS.

DR PIR: A40244; GNRVTV.

DR PDB: IN64; 25-FEB-03.

DR PDB: INS3; 08-APR-98.

DR MEROPS: S29.001; ..

DR MEROPS: U39.001; ..

DR InterPro: IPR001410; DEAD.

DR InterPro: IPR002522; HCV_capsid.

DR InterPro: IPR002521; HCV_core.

DR InterPro: IPR002519; HCV_env.

DR InterPro: IPR002531; HCV_NS1.

DR InterPro: IPR002518; HCV_NS2.

DR InterPro: IPR004109; HCV_NS3.

DR InterPro: IPR000745; HCV_NS4a.

DR InterPro: IPR001490; HCV_NS4b.

DR InterPro: IPR002868; HCV_NS5a.

DR InterPro: IPR002166; HCV_NS5b.

DR InterPro: IPR007095; RNA_pol_DS_PS.

DR InterPro: IPR007094; RNA_pol_PSVir.

DR Pfam: PF01543; HCV_capsid; 1.

DR Pfam: PF01542; HCV_core; 1.

DR Pfam: PF01539; HCV_env; 1.

DR Pfam: PF01560; HCV_NS1; 1.

DR Pfam: PF01538; HCV_NS2; 1.

DR Pfam: PF02907; HCV_NS3; 1.

DR Pfam: PF01006; HCV_NS4a; 1.

DR Pfam: PF01001; HCV_NS4b; 1.

DR Pfam: PF01506; HCV_NS5a; 1.

DR Pfam: PF00271; helicase_C; 1.

DR Pfam: PF00098; Viral_RDRP; 1.

DR ProDom: PD185062; HCV_NS1; 1.

DR SMART: SM00487; DEXdc; 1.

KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;

KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;

KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;

KW 3D-structure.

FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE

FT CELLULAR AMINOPEPTIDASE.

FT CORE PROTEIN (POTENTIAL).

FT CHAIN 1 115

FT CHAIN 116 191

FT CHAIN 192 383

FT CHAIN 384 729

FT CHAIN 730 1006

FT CHAIN 1007 1615

FT CHAIN 1616 1862

FT CHAIN 1863 2013

FT CHAIN 2014 3010

FT TRANSMEM 347 369

FT ACT_SITE 1083 1083

FT ACT_SITE 1107 1107

FT ACT_SITE 1165 1165

FT CHARGE_RELAY 1230 1237

FT NP_BIND 1316 1319

FT SITE 1316 1319

FT DECH_BOX.

FT CARBOHYD 196 196

FT CARBOHYD 209 209

FT CARBOHYD 233 233

FT CARBOHYD 234 234

FT CARBOHYD 250 250

FT CARBOHYD 250 250

FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 3010 AA: 327047 MW: 542675 D5CDFE215 CRC64;

Query Match 98.0%; Score 48; DB 1; Length 3010;

Best Local Similarity 90.9%; Pred. No. 1.1;

Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GSVVIVGRIVL 11

Db 1678 GSVVIVGRIVL 1688

RESULT 5

POLG_HCVH STANDARD; PRT: 3011 AA.

ID POLG_HCVH

AC P27958;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);

DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2

DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)

DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)

DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein

DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein

DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].

OS Hepatitis C virus (isolate H) (HCV).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OX NCBI_TaxID=11108;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=92052256; PubMed=1658800;

RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,

RA Prince A.M.;

RT "Genomic structure of the human prototype strain H of hepatitis C

RT virus: comparison with American and Japanese isolates.";

RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).

RN [2]

RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.

RX MEDLINE=97331322; PubMed=9187654;

RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;

RT "Structure of the hepatitis C virus RNA helicase domain.";

RL Nat. Struct. Biol. 4:463-467(1997).

RN [3]

RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.

RX MEDLINE=98154321; PubMed=9493270;

RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,

RA Murcko M.A., Lin C., Caron P.R.;

RT "Hepatitis C virus NS3 RNA helicase domain with a bound

RT oligonucleotide: the crystal structure provides insights into the mode

RT of unwinding.";

RL Structure 6:89-100(1998).

CC -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.

CC -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF

CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B

CC -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE

CC ACTIVATION OF NS3.

QY 1 GSVWVGRVIL 11
 DB 1678 GCVWVGRVIL 1688
 RESULT 7
 CTPC_MYCTU STANDARD; PRT: 718 AA.
 AC P96875;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Probable cation-transporting P-type ATPase C (EC 3.6.3.-).
 GN CTPC OR RV3270 OR MT3370 OR MYCY71.10.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 Gordon S.V., Eigmler K., Gas S., Barry C.E. III, Tekait F.,
 Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence.";
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 laboratory strains.";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- CATALYTIC ACTIVITY: ATP + H(2)O -> ADP + phosphate.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- SIMILARITY: Belongs to the cation transport ATPases family (P-type
 ATPases). Subfamily IB.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Z92771; CAB07083.1; -;
 CC EMBL: AE007146; AAK47711.1; -;
 CC F1R: G70978; G70978.
 CC TIGR: MT3370; -;
 CC TubercuList: Rv3270; -;
 CC InterPro: IPR006416; ATPase-IB_hvy.
 CC InterPro: IPR001757; ATPase-El-E2.
 CC InterPro: IPR006404; Heavy_metal_ATPase.
 CC InterPro: IPR005834; Hydrolyase.
 CC Pfam: PF00122; El-E2_ATPase; 1.
 CC Pfam: PF00702; Hydrolyase; 1.
 CC PRINTS: PR00119; CATATPASE.
 CC TIGRFAMs: TIGR01512; ATPase-IB2_Cd; 1.
 CC TIGRFAMs: TIGR01525; ATPase-IB_hvy; 1.
 CC TIGRFAMs: TIGR01494; ATPase_P-type; 3.
 CC PROSITE: PS00154; ATPASE_E1_E2; 1.

KW Hydrolase; Transmembrane; Phosphorylation; Magnesium; ATP-binding;
 KW Complete proteome.
 KW TRANSMEM 174 194 POTENTIAL.
 FT TRANSMEM 360 380 POTENTIAL.
 FT TRANSMEM 616 636 POTENTIAL.
 FT TRANSMEM 677 697 POTENTIAL.
 FT MOD_RES 408 408 PHOSPHORYLATION (BY SIMILARITY).
 FT METAL 610 610 MAGNESIUM (BY SIMILARITY).
 FT METAL 614 614 MAGNESIUM (BY SIMILARITY).
 SQ SEQUENCE 718 AA: 76495 MW: 85D6C93AFE636315 CRC64;
 Query Match 75.5%; Score 37; DB 1; Length 718;
 Best Local Similarity 63.6%; Pred. NO. 30;
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 QY 1 GSVWVGRVIL 11
 DB 286 GSVWVGRVIL 296
 RESULT 8
 DP02_SCHPO STANDARD; PRT: 574 AA.
 ID DP02_SCHPO
 AC 074946;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable DNA polymerase alpha subunit B.
 GN SPCC553.09C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeall C.,
 Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 Woodward J., Volkart G., Aert R., Robben J., Grynoprez B.,
 Weijtjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
 Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 Lucas M., Rochet M., Gallard C., Tallada V.A., Garzon A., Thode G.,
 Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomyces pombe.";
 RL Nature 415:871-880(2002).
 CC -!- FUNCTION: MAY PLAY AN ESSENTIAL ROLE AT THE EARLY STAGE OF
 CC CHROMOSOMAL DNA REPLICATION BY COUPLING THE POLYMERASE
 CC ALPHA/PRIMASE COMPLEX TO THE CELLULAR REPLICATION MACHINERY (BY
 CC SIMILARITY).
 CC -!- SUBUNIT: DNA POLYMERASE ALPHA-PRIMASE IS A FOUR SUBUNIT ENZYME
 CC (SUBUNITS A, B, C AND D), WHICH IS ASSEMBLED THROUGHOUT THE CELL
 CC CYCLE. THE LARGEST SUBUNIT (SUBUNIT A) HAS DNA POLYMERASE
 CC ACTIVITY, THE TWO SMALLEST SUBUNITS (SUBUNITS C AND D) HAVE DNA
 CC PRIMASE ACTIVITY. SUBUNIT B BINDS TO SUBUNIT A.

CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- PTM: PHOSPHORYLATED IN A CELL CYCLE-DEPENDENT MANNER (BY
 CC -!- SIMILARITY: BELONGS TO THE DNA POLYMERASE ALPHA SUBUNIT B FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AL023704; CAA19261.1; -.
 CC PIR; T41395; T41395.
 CC DR GeneDB_Spombe; SPCC553.09c; -.
 CC DR Pfam; PF04058; DNA_pol_alpha_B; 1.
 CC KW DNA replication; Nuclear protein; Phosphorylation.
 CC SQ SEQUENCE 574 AA; 64184 MW; 18C17FCFA7AA098A CRC64;
 CC -----
 CC Query Match 71.4%; Score 35; DB 1; Length 574;
 CC Best Local Similarity 77.8%; Pred. No. 58;
 CC Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 CC -----
 CC QY 3 VVIVGRIVL 11
 CC ||:|||||
 CC Db 205 VVVVGRIVV 213
 CC -----
 CC RESULT 9
 CC ID Y007_YEAST STANDARD; PRT; 196 AA.
 CC AC Q04487;
 CC DT 01-NOV-1997 (Rel. 35, Created)
 CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
 CC DT 15-SEP-2003 (Rel. 42, Last annotation update)
 CC DE Putative succinate dehydrogenase cytochrome B subunit, mitochondrial
 CC precursor.
 CC GN YMR118C OR YMR718.17C.
 CC OS Saccharomyces cerevisiae (Baker's yeast).
 CC OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 CC OX NCBI_TaxID=4932;
 CC RN SEQUENCE FROM N.A.
 CC RC STRAIN=S288c / AB972;
 CC RX PubMed=9169872;
 CC RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
 CC Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
 CC Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,
 CC Rice P., Skelton J., Walsh S., Whitehead S., Barrrell B.G.;
 CC "The nucleotide sequence of Saccharomyces cerevisiae chromosome
 CC XIII.";
 CC RL Nature 387:90-93(1997).
 CC CC -!- FUNCTION: MONO-HEME CYTOCHROME B. INVOLVED IN SYSTEM II OF THE
 CC MITOCHONDRIAL ELECTRON TRANSPORT CHAIN WHICH IS RESPONSIBLE FOR
 CC TRANSFERRING ELECTRONS FROM SUCCINATE TO UBIQUINONE (COENZYME Q)
 CC (BY SIMILARITY).
 CC CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
 CC inner membrane (By similarity).
 CC CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME B560 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Z49702; CAA89756.1; -.
 CC DR PIR; S54580; S54580.
 CC SGD; S0004724; YMR118C.

DR InterPro; IPR000701; Sdh_cyt.
 DR Pfam; PF01127; Sdh_cyt; 1.
 DR PROSITE; PS01000; SDH_CYT_1; 1.
 DR PROSITE; PS01001; SDH_CYT_2; 1.
 KW Hypothetical protein; Tricarboxylic acid cycle; Electron transport;
 KW Heme; Transmembrane; Mitochondrion; Transit peptide.
 FT TRANSIT ? 196 MITOCHONDRION (POTENTIAL).
 FT CHAIN 1 196 PUTATIVE SUCCINATE DEHYDROGENASE
 FT CYTOCHROME B SUBUNIT.
 FT TRANSMEM 99 119 POTENTIAL.
 FT TRANSMEM 175 195 POTENTIAL.
 FT SEQUENCE 196 AA; 22309 MW; 41413998B9B2B057 CRC64;
 CC -----
 CC Query Match 69.4%; Score 34; DB 1; Length 196;
 CC Best Local Similarity 54.5%; Pred. No. 34;
 CC Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 CC -----
 CC QY 1 GSVVIVGRIVL 11
 CC ||:|:|:|:|
 CC Db 179 GSVVIVGRIVL 189
 CC -----
 CC RESULT 10
 CC ID MCH_METTR STANDARD; PRT; 260 AA.
 CC AC Q9RPD4;
 CC DT 28-FEB-2003 (Rel. 41, Created)
 CC DT 28-FEB-2003 (Rel. 41, Last sequence update)
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
 CC DE N(5),N(1b)-methenyltetrahydromethanopterin cyclohydrolase
 CC (EC 3.5.4.27) (Methenyl-H4MPT cyclohydrolase) (Fragment).
 CC GN MCH.
 CC OS Methylosinus trichosporium.
 CC OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 CC OC Methylocystaceae; Methylosinus.
 CC OX NCBI_TaxID=426;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=OB3b;
 CC RX MEDLINE=99412275; PubMed=10482517;
 CC RA Vorholt J.A., Chistoserdova L.V., Stolyar S.M., Thauer R.K.,
 CC Lidstrom M.E.;
 CC "Distribution of tetrahydromethanopterin-dependent enzymes in
 CC methylophilic bacteria and phylogeny of methenyl
 CC tetrahydromethanopterin cyclohydrolases.";
 CC RL J. Bacteriol. 181:5750-5757(1999).
 CC CC -!- FUNCTION: Catalyzes the hydrolysis of methenyl-H(4)MPT to N(5)-
 CC formyl-H(4)MPT (By similarity).
 CC CC -!- CATALYTIC ACTIVITY: 5,10-methenyl-5,6,7,8-tetrahydromethanopterin
 CC + H(2)O = N(5)-formyl-5,6,7,8-tetrahydromethanopterin.
 CC CC -!- PATHWAY: H(4)MPT-dependent pathway of formaldehyde oxidation;
 CC third step.
 CC CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC CC -!- SIMILARITY: BELONGS TO THE MCH FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF162786; AAD56174.1; -.
 CC DR HSSP; P94954; IQLM.
 CC DR HAMAP; MF_00486; -; 1.
 CC DR InterPro; IPR003209; Cyclohydrolase.
 CC DR Pfam; PF02289; MCH; 1.
 CC DR Probom; PD011637; Cyclohydrolase; 1.
 CC KW Hydrolase; One-carbon metabolism.
 CC FT NON_TER 1
 CC SQ SEQUENCE 260 AA; 27622 MW; BD647C0DBF03C6A8 CRC64;

Query Match 69.4% Score 34: DB 1: Length 260;
 Best Local Similarity 60.0% Pred. No. 43;
 Matches 6: Conservative 3: Mismatches 1: Indels 0: Gaps 0;

Qy 1 GSVVIVGR1V 10
 /// :|||:
 Db 121 GSVVQVGRVL 130

RESULT 11

ENV_EIAV1 STANDARD; PRT: 859 AA.
 AC P22427;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ENV polypeptide precursor (Coat polypeptide) [Contains: Coat protein
 DE GP90; Coat protein GP45].
 GN ENV.
 OS Equine infectious anemia virus (clone P3.2-1) (EIAV).
 OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=8072070; PubMed=2825406;
 RA Payne S.L., Fang F.D., Liu C.P., Dhruva B.R., Rambo P., Issel C.J.,
 RA Montelaro R.C.;
 RT "Antigenic variation and lentivirus persistence: variations in
 RT envelope gene sequences during EIAV infection resemble changes
 RT reported for sequential isolates of HIV.";
 RL Virology 161:321-331(1987).
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

EMBL: M18385; AAA66407.1;
 PIR: A34027; VCLJE1.
 InterPro: IPR001027; Gp45_EIAV.
 DR InterPro: IPR001361; Gp90_EIAV.
 DR Pfam: PF01045; EIAV_GP45; 1.
 DR Pfam: PF00971; EIAV_GP90; 1.
 DR Coat protein; Glycoprotein; Polyprotein; Transmembrane; Signal.
 FT SIGNAL 1 22
 FT CHAIN 23 859 ENV POLYPEPTIDE.
 FT CHAIN 23 444 COAT PROTEIN GP90.
 FT CHAIN 445 859 COAT PROTEIN GP45.
 FT TRANSMEM 75 93 POTENTIAL.
 FT TRANSMEM 446 462 POTENTIAL.
 FT TRANSMEM 614 636 POTENTIAL.
 FT TRANSMEM 787 807 POTENTIAL.
 FT TRANSMEM 816 835 POTENTIAL.
 FT CARBOHYD 40 40 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 214 214 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 244 244 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 340 340 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 368 368 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 399 399 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 406 406 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 411 411 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 422 422 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 490 490 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 550 550 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 557 557 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 752 752 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 859 AA; 97140 MW; F4A0C071396DA867 CRC64;

Query Match 69.4% Score 34: DB 1: Length 859;
 Best Local Similarity 77.8% Pred. No. 1.3e+02;
 Matches 7: Conservative 1: Mismatches 1: Indels 0: Gaps 0;

Qy 1 GSVVIVGR1 9
 | :|||:
 Db 801 GLVIVGR1 809

RESULT 12

ENV_EIAV2 STANDARD; PRT: 859 AA.
 AC P22428;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ENV polypeptide precursor (Coat polypeptide) [Contains: Coat protein
 DE GP90; Coat protein GP45].
 GN ENV.
 OS Equine infectious anemia virus (clone P3.2-2) (EIAV).
 OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11667;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=8072070; PubMed=2825406;
 RA Payne S.L., Fang F.D., Liu C.P., Dhruva B.R., Rambo P., Issel C.J.,
 RA Montelaro R.C.;
 RT "Antigenic variation and lentivirus persistence: variations in
 RT envelope gene sequences during EIAV infection resemble changes
 RT reported for sequential isolates of HIV.";
 RL Virology 161:321-331(1987).
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

EMBL: M18386; AAA66408.1;
 PIR: B34027; VCLJE2.
 InterPro: IPR001027; Gp45_EIAV.
 DR InterPro: IPR001361; Gp90_EIAV.
 DR Pfam: PF01045; EIAV_GP45; 1.
 DR Pfam: PF00971; EIAV_GP90; 1.
 DR Coat protein; Glycoprotein; Polyprotein; Transmembrane; Signal.
 FT SIGNAL 1 22
 FT CHAIN 23 859 ENV POLYPEPTIDE.
 FT CHAIN 23 444 COAT PROTEIN GP90.
 FT CHAIN 445 859 COAT PROTEIN GP45.
 FT TRANSMEM 75 93 POTENTIAL.
 FT TRANSMEM 446 462 POTENTIAL.
 FT TRANSMEM 614 636 POTENTIAL.
 FT TRANSMEM 787 807 POTENTIAL.
 FT TRANSMEM 816 835 POTENTIAL.
 FT CARBOHYD 40 40 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 201 201 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 214 214 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 244 244 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 282 282 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 313 313 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 340 340 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 346 346 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 368 368 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 399 399 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 406 406 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 411 411 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 422 422 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 490 490 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 550 550 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 557 557 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 752 752 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 859 AA; 97188 MW; D8624E171E39B32 CRC64;
 Query Match 69.4%; Score 34; DB 1; Length 859;
 Best Local Similarity 77.8%; Pred. No. 1.3e-02;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GSVVIVGRI 9
 Db 801 GLVIVGRI 809

RESULT 13
 ENV_EIAV3 STANDARD; PRT; 859 AA.
 AC P22429;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ENV polypeptide precursor (Coat polypeptide) [Contains: Coat protein
 GP90; Coat protein GP45].
 GN ENV.
 OS Equine infectious anemia virus (clone P3.2-3) (EIAV).
 OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11668;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=8807207; PubMed=2825406;
 RA Payne S.L., Fang F.D., Liu C.P., Dhruva B.R., Rambo P., Issel C.J.,
 Montelaro R.C.;
 RT "Antigenic variation and lentivirus persistence: variations in
 envelope gene sequences during EIAV infection resemble changes
 reported for sequential isolates of HIV.";
 RL Virology 161:321-331(1987).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----

EMBL; M18387; AAA66409.1; -
 PIR; C34027; VCLJ23.
 DR InterPro: IPR001027; GP45_EIAV.
 DR InterPro: IPR001361; GP90_EIAV.
 DR Pfam; PF01045; EIAV_GP45; 1.
 DR Pfam; PF00971; EIAV_GP90; 1.
 KW Coat protein; Glycoprotein; Polypeptide; Transmembrane; Signal.
 FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 859 ENV POLYPEPTIDE.
 FT CHAIN 23 444 COAT PROTEIN GP90.
 FT CHAIN 445 859 COAT PROTEIN GP45.
 FT CHAIN 75 93 POTENTIAL.
 FT TRANSMEM 446 462 POTENTIAL.
 FT TRANSMEM 614 636 POTENTIAL.
 FT TRANSMEM 787 807 POTENTIAL.
 FT TRANSMEM 816 835 POTENTIAL.
 FT CARBOHYD 40 40 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 214 214 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 244 244 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 340 340 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 368 368 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 399 399 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 406 406 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 411 411 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 422 422 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 490 490 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 550 550 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 557 557 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 752 752 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 859 AA; 97066 MW; 982A9F5A1AD8FA4D CRC64;

Query Match 69.4%; Score 34; DB 1; Length 859;
 Best Local Similarity 77.8%; Pred. No. 1.3e-02;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GSVVIVGRI 9
 Db 801 GLVIVGRI 809

RESULT 14
 ENV_EIAV9 STANDARD; PRT; 859 AA.
 AC P11306;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ENV polypeptide precursor (Coat polypeptide) [Contains: Coat protein
 GP90; Coat protein GP45].
 GN ENV.
 OS Equine infectious anemia virus (clone 1369) (EIAV).
 OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11670;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87236196; PubMed=3035786;
 RA Kawakami T., Sherman L., Dahlberg J., Gazit A., Yaniv A.,
 Tronick S.R., Aaronson S.A.;
 RT "Nucleotide sequence analysis of equine infectious anemia virus
 proviral DNA.";
 RL Virology 158:300-312(1987).
 RN [2]
 RP REVISIONS TO N-TERMINUS.
 RA Tronick S.R.;
 RL Submitted (NOV-1987) to the EMBL/GenBank/DBJ databases.
 CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----

EMBL; M16575; AAB59863.1; -
 DR InterPro: IPR001027; GP45_EIAV.
 DR InterPro: IPR001361; GP90_EIAV.
 DR Pfam; PF01045; EIAV_GP45; 1.
 DR Pfam; PF00971; EIAV_GP90; 1.
 KW Coat protein; Glycoprotein; Polypeptide; Transmembrane; Signal.
 FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 859 ENV POLYPEPTIDE.
 FT CHAIN 23 444 COAT PROTEIN GP90.
 FT CHAIN 445 859 COAT PROTEIN GP45.
 FT TRANSMEM 75 93 POTENTIAL.
 FT TRANSMEM 446 462 POTENTIAL.
 FT TRANSMEM 614 636 POTENTIAL.
 FT TRANSMEM 787 807 POTENTIAL.

FT TRANSMEM 816 835 POTENTIAL.
 FT CARBOHYD 40 40 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 214 214 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 244 244 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 340 340 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 368 368 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 399 399 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 411 411 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 490 490 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 550 550 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 557 557 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 752 752 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 859 AA; 97113 MW; 484ED8518CDAF364 CRC64;

Query Match 69.4%; Score 34; DB 1; Length 859;
 Best Local Similarity 77.8%; Pred. No. 1.3e+02;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GSWIVIGRI 9

Db 801 GLVIVIGRI 809

RESULT 15

ENV_EIACV

ID ENV_EIACV STANDARD; PRT; 859 AA.
 AC P32541;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ENV polyprotein precursor (Coat polyprotein) [Contains: Coat protein
 DE GP90; Coat protein GP45].
 GN ENV.
 OS Equine infectious anemia virus (clone CL22) (EIAV).
 OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
 OX NCBI_TaxID=31675;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9229230; PubMed=1318398;
 RA Perry S.T., Flaherty M.T., Kelley M.J., Clabough D.L., Tronick S.R.,
 RA Coggins L., Whetter L., Lengel C.R., Fuller F.;
 RT "The surface envelope protein gene region of equine infectious anemia
 RT virus is not an important determinant of tropism in vitro.";
 RL J. Virol. 66:4085-4097(1992).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed, usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: M87581; AAA43005.1; -
 DR PIR: C41991; VCLJ22.
 DR InterPro: IPR001027; Gp45_EIACV.
 DR InterPro: IPR001361; Gp90_EIACV.
 DR Pfam: PF01045; EIAV_GP45_1.
 DR Pfam: PF00971; EIAV_GP90_1.
 KW Coat protein; Glycoprotein; Polyprotein; Transmembrane; Signal.

FT SIGNAL 1 22 POTENTIAL.

FT CHAIN 23 859 ENV POLYPROTEIN.

FT CHAIN 23 444 COAT PROTEIN GP90.

FT CHAIN 445 859 COAT PROTEIN GP45.

FT TRANSMEM 75 93 POTENTIAL.

FT TRANSMEM 446 472 POTENTIAL.

FT TRANSMEM 617 636 POTENTIAL.
 FT TRANSMEM 787 807 POTENTIAL.
 FT CARBOHYD 816 835 POTENTIAL.
 FT CARBOHYD 40 40 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 214 214 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 244 244 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 340 340 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 368 368 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 399 399 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 406 406 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 411 411 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 490 490 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 550 550 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 557 557 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 752 752 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 859 AA; 97140 MW; 23E020E80DF334FA CRC64;

Query Match 69.4%; Score 34; DB 1; Length 859;
 Best Local Similarity 77.8%; Pred. No. 1.3e+02;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GSWIVIGRI 9

Db 801 GLVIVIGRI 809

Search completed: August 30, 2003, 19:13:51

Job time : 1.54479 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 19:00:22 ; Search time 2.09905 seconds

(without alignments)
1352.314 Million cell updates/sec

Title: US-09-965-594-26

Perfect score: 49

Sequence: 1 GSVVIGRVL 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rudent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49	100.0	3010	12 Q9QP06	Q9qp06 hepatitis c
2	49	100.0	3010	12 Q9QP61	Q9qp61 hepatitis c
3	49	100.0	3010	12 Q81760	Q81760 hepatitis c
4	49	100.0	3010	12 Q9DTE5	Q9dte5 hepatitis c
5	49	100.0	4040	12 Q9IFH8	Q9ifh8 mucosal dis
6	48	98.0	88	12 Q39914	Q39914 hepatitis c
7	48	98.0	89	12 Q39895	Q39895 hepatitis c
8	48	98.0	102	12 Q9PXP5	Q9pxp5 hepatitis c
9	48	98.0	138	12 Q68209	Q68209 hepatitis c
10	48	98.0	138	12 Q68235	Q68235 hepatitis c
11	48	98.0	138	12 Q68218	Q68218 hepatitis c
12	48	98.0	138	12 Q68244	Q68244 hepatitis c
13	48	98.0	138	12 Q68206	Q68206 hepatitis c
14	48	98.0	138	12 Q68242	Q68242 hepatitis c
15	48	98.0	138	12 Q68240	Q68240 hepatitis c
16	48	98.0	138	12 Q68213	Q68213 hepatitis c

17	48	98.0	138	12 Q68210	Q68210 hepatitis c
18	48	98.0	138	12 Q68227	Q68227 hepatitis c
19	48	98.0	138	12 Q68229	Q68229 hepatitis c
20	48	98.0	138	12 Q68207	Q68207 hepatitis c
21	48	98.0	138	12 Q68216	Q68216 hepatitis c
22	48	98.0	138	12 Q68228	Q68228 hepatitis c
23	48	98.0	138	12 Q68221	Q68221 hepatitis c
24	48	98.0	138	12 Q68205	Q68205 hepatitis c
25	48	98.0	138	12 Q68215	Q68215 hepatitis c
26	48	98.0	138	12 Q68232	Q68232 hepatitis c
27	48	98.0	138	12 Q68208	Q68208 hepatitis c
28	48	98.0	138	12 Q68231	Q68231 hepatitis c
29	48	98.0	138	12 Q68211	Q68211 hepatitis c
30	48	98.0	138	12 Q68238	Q68238 hepatitis c
31	48	98.0	138	12 Q68237	Q68237 hepatitis c
32	48	98.0	138	12 Q68217	Q68217 hepatitis c
33	48	98.0	138	12 Q68230	Q68230 hepatitis c
34	48	98.0	172	12 Q81579	Q81579 hepatitis c
35	48	98.0	172	12 Q81577	Q81577 hepatitis c
36	48	98.0	172	12 Q81575	Q81575 hepatitis c
37	48	98.0	172	12 Q81582	Q81582 hepatitis c
38	48	98.0	172	12 Q81584	Q81584 hepatitis c
39	48	98.0	172	12 Q81574	Q81574 hepatitis c
40	48	98.0	172	12 Q81578	Q81578 hepatitis c
41	48	98.0	172	12 Q81583	Q81583 hepatitis c
42	48	98.0	172	12 Q81581	Q81581 hepatitis c
43	48	98.0	271	12 Q81573	Q81573 hepatitis c
44	48	98.0	425	12 Q68344	Q68344 hepatitis c
45	48	98.0	1186	12 Q81755	Q81755 hepatitis c

ALIGNMENTS

RESULT 1

Q9QP06 PRELIMINARY; PRT: 3010 AA.

ID Q9QP06; AC Q9QP06; DT 01-MAY-2000 (Tremblrel. 13, Created)

DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)

DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)

DE Genome polyprotein.

OS Hepatitis C virus type 1b.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OX NCBI_TaxID=31647;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-RB;

RA Bartschlagler R.;

RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-RB;

RX MEDLINE=99370154; PubMed=10438800;

RA Koch J.O., Bartschlagler R.;

RT "Modulation of hepatitis C virus NS5A hyperphosphorylation by nonstructural proteins NS3, NS4A, and NS4B.";

RL J. Virol. 73:7138-7146(1999).

CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA (BY SIMILARITY).

CC EMBL; AJ238800; CAB53095.1; -.

DR HSSP; P26663; IN63.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002522; HCV_capsid.

DR InterPro; IPR002521; HCV_core.

DR InterPro; IPR002519; HCV_env.

DR InterPro; IPR002531; HCV_NS1.

DR InterPro; IPR002518; HCV_NS2.

DR InterPro; IPR004109; HCV_NS3.

DR InterPro; IPR000745; HCV_NS4a.

DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RDRP.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00998; Viral_RDRP; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR PROSITE: PS05057; RDRP_POSITIVE; 1.
 DR PROSITE: PS0521; RDRP_VIRAL; 1.
 DR Coats protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
 FT CHAIN 1 191
 FT CHAIN 192 383
 FT CHAIN 384 746
 FT CHAIN 747 809
 FT CHAIN 810 1026
 FT CHAIN 1027 1657
 FT CHAIN 1658 1711
 FT CHAIN 1712 1572
 FT CHAIN 1973 2418
 FT CHAIN 2419 3010
 SQ SEQUENCE 3010 AA; 336999 MW; A570B980DD64634 CRC64;

Query Match 100.0%; Score 49; DB 12; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 6.6;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSVVIVGRVL 11
 |||||
 Db 1678 GSVVIVGRVL 1688

RESULT 2
 Q90P61 PRELIMINARY: PRT: 3010 AA.
 AC Q90P61:
 DT 01-MAY-2000 (TRENBLrel. 13, Created)
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE Genome polyprotein.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RC STRAIN=HC-C2;
 RX MEDLINE=93359897; PubMed=8394876;
 RA Wang Y., Okamoto H., Tsuda F., Nagayama K., Tao Q.M., Mishiro S.;
 RT "Prevalence, Genotypes, and an isolate(HC-C2) of Hepatitis C Virus in
 RA Chinese Patients With Liver Disease.";
 RL J. Med. Virol. 40:254-260(1993).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 CC EMBL: AF176573; AAD50312.1; -;
 DR HSP: P26663; INS3.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR002531; HCV_NS1.

DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RDRP.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00998; Viral_RDRP; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR PROSITE: PS05057; RDRP_POSITIVE; 1.
 DR PROSITE: PS0521; RDRP_VIRAL; 1.
 DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
 KW Hydrolyase; Nonstructural protein; Polyprotein;
 KW RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3010 AA; 327068 MW; 9105F69483DD5BBA CRC64;

Query Match 100.0%; Score 49; DB 12; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 6.6;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSVVIVGRVL 11
 |||||
 Db 1678 GSVVIVGRVL 1688

RESULT 3
 Q81760 PRELIMINARY: PRT: 3010 AA.
 AC Q81760:
 DT 01-NOV-1996 (TRENBLrel. 01, Created)
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE Genome polyprotein.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=HC-C2;
 RX MEDLINE=93359897; PubMed=8394876;
 RA Wang Y., Okamoto H., Tsuda F., Nagayama K., Tao Q.M., Mishiro S.;
 RT "Prevalence, Genotypes, and an isolate(HC-C2) of Hepatitis C Virus in
 RT Chinese Patients With Liver Disease.";
 RL J. Med. Virol. 40:254-260(1993).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 CC EMBL: D10934; BAA01728.1; -;
 DR HSP: P26663; INS3.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.

DR	InterPro: IPR002868; HCV_NS5a.
DR	InterPro: IPR002166; HCV_RdRP.
DR	InterPro: IPR007095; RNA_pol_DS_PS.
DR	InterPro: IPR007094; RNA_pol_PSVlr.
DR	Pfam: PF01543; HCV_capsid; 1.
DR	Pfam: PF01542; HCV_core; 1.
DR	Pfam: PF01539; HCV_env; 1.
DR	Pfam: PF01560; HCV_NSI; 1.
DR	Pfam: PF01538; HCV_NS2; 1.
DR	Pfam: PF02907; HCV_NS3; 1.
DR	Pfam: PF01006; HCV_NS4a; 1.
DR	Pfam: PF01001; HCV_NS4b; 1.
DR	Pfam: PF01506; HCV_NS5a; 1.
DR	Pfam: PF00998; Viral_RdRP; 1.
DR	ProDom: PD186062; HCV_NSI; 1.
DR	SMART: SM00487; DEXDC; 1.
DR	DR PROSITE: PS05057; RDRP_POSITIVE; 1.
DR	DR PROSITE: PS0521; RDRP_VIRAL; 1.
KW	Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW	Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SEQUENCE	3010 AA; 326855 HW; EA7D306A4BA2E224 CRC64;
Query Match	100.0%; Score 49; DB 12; Length 3010;
Best Local Similarity	100.0%; Pred. No. 6.6;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 GSVWVGRIVL 11
DB	1678 GSVWVGRIVL 1688
RESULT 4	
Q9DTE5	PRELIMINARY: PRT; 3010 AA.
ID	Q9DTE5
AC	Q9DTE5
DT	01-MAR-2001 (TREMBLrel. 16, Created)
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE	Genome polyprotein.
OS	Hepatitis C virus.
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC	Hepacivirus.
OX	NCBI_TaxID=11103;
RN	{1}
RP	SEQUENCE FROM N.A.
RC	STRAIN=HCV145;
RA	Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,
RA	Hatahara T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,
RA	Mishiro S.;
RT	"Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients
RT	with hepatocellular carcinoma: the progression score revisited.;"
RL	Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
CC	-1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC	LIPIDPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC	PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC	PROTEIN C AND MRNA (BY SIMILARITY).
DR	EMBL: AB049092; BAB18805.1; ..
DR	HSP: P27958; 1HEI;
DR	InterPro: IPR000345; CytC_heme_bind.
DR	InterPro: IPR001410; DEAD.
DR	InterPro: IPR002522; HCV_capsid.
DR	InterPro: IPR002521; HCV_core.
DR	InterPro: IPR002519; HCV_env.
DR	InterPro: IPR002531; HCV_NSI.
DR	InterPro: IPR002518; HCV_NS2.
DR	InterPro: IPR004109; HCV_NS3.
DR	InterPro: IPR00045; HCV_NS4a.
DR	InterPro: IPR001490; HCV_NS4b.
DR	InterPro: IPR002868; HCV_NS5a.
DR	InterPro: IPR002166; HCV_RdRP.
DR	InterPro: IPR001650; Helicase.C.
DR	InterPro: IPR007095; RNA_pol_DS_PS.
DR	InterPro: IPR007094; RNA_pol_PSVlr

DR PROSITE; P50507; RDRP_POSITIVE; 1.
DR PROSITE; P50521; RDRP_VIRAL; 1.
DR PROSITE; P50531; RNASE_L2; 1.
KW ATP-binding; Helicase; Hydrolase; Nonstructural protein; Polyprotein;
SQ RNA-directed RNA polymerase; Transferase.
SQ SEQUENCE 4040 AA; 453073 MW; ADE87791D055B9DC CRC64;

Query Match 100.0%; Score 49; DB 12; Length 4040;
Best Local Similarity 100.0%; Pred. No. 8.8;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GSVVIVGRIVL 11
| | | | | | | | | |
Db 10 GSVVIVGRIVL 20

RESULT 6
O39914 PRELIMINARY; PRT; 88 AA.
AC O39914;
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Non-structural protein 4a/b (Fragment).
GN NS4A/B.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FO1;
RX MEDLINE=98032593; PubMed=9365889;
RA Prescott L.E., Berger A., Pawlotsky J.M., Conjeevaram P., Pike I.,
RA Simmonds P.;
RT *Significance analysis of hepatitis C virus variants producing discrepant
RT results with two different genotyping assays.*;
RL J. Med. Virol. 53:237-244(1997).
DR EMBL; AF007519; AAB62970.2; -.
DR InterPro; IPR000745; HCV_NS4a.
DR Pfam; PF01006; HCV_NS4a; 1.
FT NON_TER 1 1
FT NON_TER 88 88
SQ SEQUENCE 88 AA; 9750 MW; BF7B5198B317B6E0 CRC64;

Query Match 98.0%; Score 48; DB 12; Length 88;
Best Local Similarity 90.9%; Pred. No. 0.3;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GSVVIVGRIVL 11
| | | | | | | | | |
Db 6 GSVVIVGRIVL 16

RESULT 7
O39895 PRELIMINARY; PRT; 89 AA.
AC O39895;
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE Non-structural protein 4a/b (Fragment).
GN NS4A/B.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Prescott L.E., Berger A., Pawlotsky J.M., Conjeevaram P., Pike I.,
RA Simmonds P.;
RT *Significance analysis of Hepatitis C virus variants producing discrepant
RT results with two different genotyping assays.*;

RL J. Med. Virol. 0:0-0(1997).
DR EMBL; AF007500; AAB62951.1; -.
DR InterPro; IPR000745; HCV_NS4a.
DR Pfam; PF01006; HCV_NS4a; 1.
FT NON_TER 1 1
FT NON_TER 89 89
SQ SEQUENCE 89 AA; 9748 MW; 581BB8C8A3EA8B5C CRC64;

Query Match 98.0%; Score 48; DB 12; Length 89;
Best Local Similarity 90.9%; Pred. No. 0.3;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GSVVIVGRIVL 11
| | | | | | | | | |
Db 8 GSVVIVGRIVL 18

RESULT 8
O9PXP5 PRELIMINARY; PRT; 102 AA.
AC O9PXP5;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Non-structural protein NS4-GROUP II HCV-specific antigen C14-1
DE (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94245087; PubMed=7514558;
RA Tanaka T., Tsukiyama-Kohara K., Yamaguchi K., Yagi S., Tanaka S.,
RA Hasegawa A., Ohta Y., Hattori N., Kohara M.;
RT *Significance of specific antibody assay for genotyping of hepatitis C
RT virus.*;
RL Hepatology 19:1347-1353(1994).
DR InterPro; IPR000745; HCV_NS4a.
DR Pfam; PF01006; HCV_NS4a; 1.
FT NON_TER 102 102
FT NON_TER 102 102
SQ SEQUENCE 102 AA; 11419 MW; 08124C19CF367F06 CRC64;

Query Match 98.0%; Score 48; DB 12; Length 102;
Best Local Similarity 90.9%; Pred. No. 0.35;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GSVVIVGRIVL 11
| | | | | | | | | |
Db 20 GSVVIVGRIVL 30

RESULT 9
Q68209 PRELIMINARY; PRT; 138 AA.
ID Q68209;
AC Q68209;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE Nonstructural protein (Fragment).
GN NS4.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1b;
RX MEDLINE=95146953; PubMed=7844535;
RA Greene W.K., Cheong M.K., Ng V., Yap K.W.;
RT *Prevalence of hepatitis C virus sequence variants in South-East
RT Asia.*;
RL J. Gen. Virol. 76:211-215(1995).

```
DR EMBL: U14245; AAC53934.1; -.
DR HSSP: P27958; IHEI.
DR InterPro: IPR000745; HCV_NS4a.
DR Pfam: PF01006; HCV_NS4a; 1.
DR NON_TER 1
FT NON_TER 138
SQ SEQUENCE 138 AA; 15149 MW; DBAE62A0FE9E2D57 CRC64;

Query Match          98.0%; Score 48; DB 12; Length 138;
Best Local Similarity 90.9%; Pred. No. 0.47;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSVVIVGRIVL 11
DB 52 GSVVIVGRIL 62

RESULT 10
Q68235 PRELIMINARY; PRT; 138 AA.
AC Q68235;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Nonstructural protein (Fragment).
GN NS4.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1b;
RX MEDLINE=95146953; PubMed=7844535;
RA Greene W.K., Cheong M.K., Ng V., Yap K.W.;
RT "Prevalence of hepatitis C virus sequence variants in South-East
RT Asia.";
RL J. Gen. Virol. 76:211-215(1995).
DR EMBL: U14271; AAC53960.1; -.
DR HSSP: P27958; IHEI.
DR InterPro: IPR000745; HCV_NS4a.
DR Pfam: PF01006; HCV_NS4a; 1.
DR NON_TER 1
FT NON_TER 138
SQ SEQUENCE 138 AA; 15104 MW; 585DC5A627D0F3E3 CRC64;

Query Match          98.0%; Score 48; DB 12; Length 138;
Best Local Similarity 90.9%; Pred. No. 0.47;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSVVIVGRIVL 11
DB 52 GSVVIVGRIL 62

RESULT 11
Q68218 PRELIMINARY; PRT; 138 AA.
AC Q68218;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Nonstructural protein (Fragment).
GN NS4.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1b;
RX MEDLINE=95146953; PubMed=7844535;
RA Greene W.K., Cheong M.K., Ng V., Yap K.W.;
```

```
RT "Prevalence of hepatitis C virus sequence variants in South-East
RT Asia.";
RL J. Gen. Virol. 76:211-215(1995).
DR EMBL: U14254; AAC53943.1; -.
DR HSSP: P27958; IHEI.
DR InterPro: IPR000745; HCV_NS4a.
DR Pfam: PF01006; HCV_NS4a; 1.
DR NON_TER 1
FT NON_TER 138
SQ SEQUENCE 138 AA; 15189 MW; DB78E92DDC67040F CRC64;

Query Match          98.0%; Score 48; DB 12; Length 138;
Best Local Similarity 90.9%; Pred. No. 0.47;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSVVIVGRIVL 11
DB 52 GSVVIVGRIL 62

RESULT 12
Q68244 PRELIMINARY; PRT; 138 AA.
AC Q68244;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Nonstructural protein (Fragment).
GN NS4.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1b;
RX MEDLINE=95146953; PubMed=7844535;
RA Greene W.K., Cheong M.K., Ng V., Yap K.W.;
RT "Prevalence of hepatitis C virus sequence variants in South-East
RT Asia.";
RL J. Gen. Virol. 76:211-215(1995).
DR EMBL: U14280; AAC53969.1; -.
DR HSSP: P27958; IHEI.
DR InterPro: IPR000745; HCV_NS4a.
DR Pfam: PF01006; HCV_NS4a; 1.
DR NON_TER 1
FT NON_TER 138
SQ SEQUENCE 138 AA; 15118 MW; B7F7EB2733770408 CRC64;

Query Match          98.0%; Score 48; DB 12; Length 138;
Best Local Similarity 90.9%; Pred. No. 0.47;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSVVIVGRIVL 11
DB 52 GSVVIVGRIL 62

RESULT 13
Q68206 PRELIMINARY; PRT; 138 AA.
AC Q68206;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Nonstructural protein (Fragment).
GN NS4.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
```


RC STRAIN=1b;
RX MEDLINE-95146953; PubMed-7844535;
RA Greene W.K., Cheong M.K., Ng V., Yap K.W.;
RT "Prevalence of hepatitis C virus sequence variants in South-East
RT Asia.";
RL J. Gen. Virol. 76:211-215(1995).
DR EMBL: U14242; AAC53931.1; -.
DR HSSP: P27958; 1HEI.
DR InterPro: IPR000745; HCV_NS4a.
DR Pfam: PF01006; HCV_NS4a; 1.
FT NON_TER 1
FT NON_TER 138 1
SQ SEQUENCE 138 AA: 15117 MW: 55953A1B74DE13E CRC64;

Query Match 98.0%; Score 48; DB 12; Length 138;
Best Local Similarity 90.9%; Pred. No. 0.47;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSVWIVGRIVL 11
Db 52 GSVWIVGRILL 62
|||||

RESULT 14

Q68242
ID Q68242 PRELIMINARY; PRT; 138 AA.
AC Q68242;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Nonstructural protein (Fragment).
GN NS4.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1b;
RX MEDLINE-95146953; PubMed-7844535;
RA Greene W.K., Cheong M.K., Ng V., Yap K.W.;
RT "Prevalence of hepatitis C virus sequence variants in South-East
RT Asia.";
RL J. Gen. Virol. 76:211-215(1995).
DR EMBL: U14278; AAC53967.1; -.
DR HSSP: P27958; 1HEI.
DR InterPro: IPR000745; HCV_NS4a.
DR Pfam: PF01006; HCV_NS4a; 1.
FT NON_TER 1
FT NON_TER 138 138
SQ SEQUENCE 138 AA: 15281 MW: CD5B5B3834C6070D CRC64;

Query Match 98.0%; Score 48; DB 12; Length 138;
Best Local Similarity 90.9%; Pred. No. 0.47;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSVWIVGRIVL 11
Db 52 GSVWIVGRILL 62
|||||

RESULT 15

Q68240
ID Q68240 PRELIMINARY; PRT; 138 AA.
AC Q68240;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Nonstructural protein (Fragment).
GN NS4.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.

OX NCBI_TaxID=111103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1b;
RX MEDLINE-95146953; PubMed-7844535;
RA Greene W.K., Cheong M.K., Ng V., Yap K.W.;
RT "Prevalence of hepatitis C virus sequence variants in South-East
RT Asia.";
RL J. Gen. Virol. 76:211-215(1995).
DR EMBL: U14276; AAC53965.1; -.
DR HSSP: P27958; 1HEI.
DR InterPro: IPR000745; HCV_NS4a.
DR Pfam: PF01006; HCV_NS4a; 1.
FT NON_TER 1
FT NON_TER 138 138
SQ SEQUENCE 138 AA: 15115 MW: 6A042677B354CA7A CRC64;

Query Match 98.0%; Score 48; DB 12; Length 138;
Best Local Similarity 90.9%; Pred. No. 0.47;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSVWIVGRIVL 11
Db 52 GSVWIVGRILL 62
|||||

Search completed: August 30, 2003, 19:18:23
Job time: 4.09905 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 30, 2003, 19:18:33 ; Search time 142.976 Seconds
(without alignments)
3147.423 Million cell updates/sec

Title: US-09-965-594-26
Perfect score: 49
Sequence: 1 GSWIVGRIVL 11

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -MODEL=frame+p2n-model -DEV=xlp
-Q=/cgn2.1/USPTO.spool/US09965594/runat_29082003.151919.28310/app_query.fasta.1.2872
-DB=GenEmbl -QFMT=fastap -SURFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.coi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09965594 -CGN=1.14686 -runat_29082003.151919.28310 -NCPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPEXT=10 -XGAPEXT=0.5 -FGAPOPT=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl :
1: gb_ba :
2: gb_htg :
3: gb_in :
4: gb_om :
5: gb_ov :
6: gb_pat :
7: gb_ph :
8: gb_pl :
9: gb_pr :
10: gb_ro :
11: gb_sts :
12: gb_sy :
13: gb_un :
14: gb_vi :
15: em_ba :
16: em_fun :
17: em_hum :
18: em_in :
19: em_mu :
20: em_om :
21: em_or :
22: em_ov :
23: em_pat :
24: em_pi :
25: em_pl :
26: em_ro :
27: em_sts :
28: em_un :

29: em_vi :
30: em_htg_hum :
31: em_htg_inv :
32: em_htg_other :
33: em_htg_mus :
34: em_htg_pln :
35: em_htg_rod :
36: em_htg_mam :
37: em_htg_vrt :
38: em_sy :
39: em_htgo_hum :
40: em_htgo_mus :
41: em_htgo_other :

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	49	100.0	618	6	E06198 E06198 cDNA encodi
2	49	100.0	618	6	E06394 E06394 cDNA encodi
3	49	100.0	7475	6	A91965 A91965 Sequence 1
4	49	100.0	7475	6	AR031992 AR031992 Sequence
5	49	100.0	7475	6	AR207294 AR207294 Sequence
6	49	100.0	9033	14	HCU238800 HCU238800 Hepatitis
7	49	100.0	9344	14	AB049092 AB049092 Hepatitis
8	49	100.0	9400	14	HPCCGENOM HPCCGENOM
9	49	100.0	9456	14	HPCRNA HPCRNA
10	49	100.0	9600	14	AF176573 AF176573 Hepatitis
11	49	100.0	12734	6	AR179057 AR179057 Sequence
12	49	100.0	12734	14	AF268278 AF268278 Hepatitis
13	48	98.0	75	6	AR145217 AR145217 Sequence
14	48	98.0	75	6	AR145221 AR145221 Sequence
15	48	98.0	78	6	AR145197 AR145197 Sequence
16	48	98.0	78	6	AR145213 AR145213 Sequence
17	48	98.0	96	6	AR145247 AR145247 Sequence
18	48	98.0	96	6	AR145249 AR145249 Sequence
19	48	98.0	161	6	AX481515 AX481515 Sequence
20	48	98.0	189	6	AR037527 AR037527 Sequence
21	48	98.0	189	6	E09290 E09290 DNA encodin
22	48	98.0	266	14	AF007519 AF007519 Hepatitis
23	48	98.0	267	6	AR037528 AR037528 Sequence
24	48	98.0	267	6	E09291 E09291 DNA encodin
25	48	98.0	267	14	AF007500 AF007500 Hepatitis
26	48	98.0	279	6	E11063 E11063 DNA sequenc
27	48	98.0	321	6	E11062 E11062 DNA sequenc
28	48	98.0	372	6	E03389 E03389 DNA encodin
29	48	98.0	414	14	HCU14241 HCU14241 Hepatitis C
30	48	98.0	414	14	HCU14242 HCU14242 Hepatitis C
31	48	98.0	414	14	HCU14243 HCU14243 Hepatitis C
32	48	98.0	414	14	HCU14244 HCU14244 Hepatitis C
33	48	98.0	414	14	HCU14245 HCU14245 Hepatitis C
34	48	98.0	414	14	HCU14246 HCU14246 Hepatitis C
35	48	98.0	414	14	HCU14247 HCU14247 Hepatitis C
36	48	98.0	414	14	HCU14249 HCU14249 Hepatitis C
37	48	98.0	414	14	HCU14251 HCU14251 Hepatitis C
38	48	98.0	414	14	HCU14252 HCU14252 Hepatitis C
39	48	98.0	414	14	HCU14253 HCU14253 Hepatitis C
40	48	98.0	414	14	HCU14254 HCU14254 Hepatitis C
41	48	98.0	414	14	HCU14257 HCU14257 Hepatitis C
42	48	98.0	414	14	HCU14263 HCU14263 Hepatitis C
43	48	98.0	414	14	HCU14264 HCU14264 Hepatitis C
44	48	98.0	414	14	HCU14265 HCU14265 Hepatitis C
45	48	98.0	414	14	HCU14266 HCU14266 Hepatitis C

ALIGNMENTS

E06198
LOCUS E06198 618 bp RNA linear PAT 29-SEP-1997
DEFINITION cDNA encoding genes derived from hepatitis C virus.
ACCESSION E06198
VERSION E06198.1 GI:2174385
KEYWORDS JP 1994000085-A/38.
SOURCE Hepatitis C virus
ORGANISM Hepatitis C virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.
REFERENCE 1 (bases 1 to 618)
AUTHORS Seki,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and Hayashi,N.
TITLE GENE OR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE CODED BY THE SAME AND ITS PRODUCTION
JOURNAL Patent: JP 1994000085-A 38 11-JAN-1994;
COMMENT MITSUBISHI KASEI CORP
OS (hepatitis C virus)
PN JP 1994000085-A/38
PD 11-JAN-1994
PF 11-JUN-1992 JP 1992194497
PR 11-JUN-1991 JP 91P 139268, 12-JUL-1991 JP 91P 172794, PR 07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR 20-APR-1992 JP 92P 99957
PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI MURAKAMI TOMOKO,
PI TERANISHI YUTAKA, HAYASHI NORIO
PC C12N15/51,C07K7/06,C07K7/10,C07K13/00,C07K15/12, PC C12N1/21,C12N5/10,
PC C12N15/11,C12N15/70,C12N15/85,C12P21/02//A61K39/29, PC (C12N1/21, (C12N1:19), (C12N5/10,C12R1:91), (C12P21/02,C12R1:19), (C12P21/02, PC C12R1:91), (C12N5/10,C12R1:91), (C12P21/02,C12R1:19), (C12P21/02, PC C07K99:00;
PC C07K99:00;
CC strandedness: Double;
CC topology: Linear;
CC anti-sense: No;
CC *source: clone-N13-1;
FH Key Location/Qualifiers
FH mat_peptide 9..608
FT /product='the peptides reacting specifically and immunochemically with the serum of hepatitis patient'.
FT type C
FT Location/Qualifiers
FT 1..618
FT /organism='Hepatitis C virus'
FT /mol_type='genomic RNA'
FT /db_xref='taxon:11103'
BASE COUNT 135 a 184 c 179 g 120 t
ORIGIN
Alignment Scores:
Pred. No.: 0.238 Length: 618
Score: 49.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-09-965-594-26 (1-11) x E06198 (1-618)
Qy 1 GlySerValValIleValGlyArgIleValLeu 11
Db 351 GCCAGCGTGCATTGTGGCAGGATCGTCTTG 383
RESULT 2
E06394
LOCUS E06394 618 bp RNA linear PAT 29-SEP-1997
DEFINITION cDNA encoding genes derived from hepatitis C virus.
ACCESSION E06394

E06394.1 GI:2174581
KEYWORDS JP 1994000086-A/38.
SOURCE Hepatitis C virus
ORGANISM Hepatitis C virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.
REFERENCE 1 (bases 1 to 618)
AUTHORS Seki,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and Hayashi,N.
TITLE GENE FOR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE CODED BY THE SAME AND ITS PRODUCTION
JOURNAL Patent: JP 1994000086-A 38 11-JAN-1994;
COMMENT MITSUBISHI KASEI CORP
OS (hepatitis C virus)
PN JP 1994000086-A/38
PD 11-JAN-1994
PF 07-OCT-1992 JP 1992293734
PR 07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR 20-APR-1992 JP 92P 99957
PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI MURAKAMI TOMOKO,
PI TERANISHI YUTAKA, HAYASHI NORIO
PC C12N15/51,C07K7/06,C07K7/10,C07K13/00,C12N5/10, PC C12N15/11,
PC C12N15/85,C12P21/02//A61K39/29, (C12P21/02,C12R1:91),C07K99:00;
CC strandedness: Double;
CC topology: Linear;
CC anti-sense: No;
CC *source: clone-N13-1;
FH Key Location/Qualifiers
FH mat_peptide 9..608
FT /product='the peptides reacting specifically and immunochemically with the serum of hepatitis patient'.
FT type C
FT Location/Qualifiers
FT 1..618
FT /organism='Hepatitis C virus'
FT /mol_type='genomic RNA'
FT /db_xref='taxon:11103'
BASE COUNT 135 a 184 c 179 g 120 t
ORIGIN
Alignment Scores:
Pred. No.: 0.238 Length: 618
Score: 49.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-09-965-594-26 (1-11) x E06394 (1-618)
Qy 1 GlySerValValIleValGlyArgIleValLeu 11
Db 351 GCCAGCGTGCATTGTGGCAGGATCGTCTTG 383
RESULT 3
A91965
LOCUS A91965 7475 bp DNA circular PAT 22-JAN-2000
DEFINITION Sequence 1 from Patent WO9822496.
ACCESSION A91965
VERSION A91965.1 GI:6740811
KEYWORDS unidentified
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 7475)
AUTHORS Attwood,M.R. and Hurst,D.N.
TITLE ANTIVIRAL PEPTIDE DERIVATIVES
JOURNAL Patent: WO 9822496-A 1 28-MAY-1998;

```

HOFFMANN LA ROCHE (CH)
FEATURES             Location/Qualifiers
  source             1..7475
                    /organism="unidentified"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:32644"
BASE COUNT          1853 a 1916 c 2039 g 1667 t
ORIGIN
Alignment Scores:      3.53      Length:      7475
Pred. No.:            49.00      Matches:      11
Score:                100.00%     Conservative: 0
Percent Similarity:    100.00%     Mismatches:  0
Best Local Similarity: 100.00%     Indels:      0
Query Match:          100.00%     Gaps:        0
DB:
US-09-965-594-26 (1-11) x A91965 (1-7475)

QY      1 GlySerValValIleValGlyArgIleValLeu 11
Db      3451 GGCAGCGTGTTCATCTGTGGCAGGATCGTCTTG 3483

RESULT 4
LOCUS      AR031992              7475 bp      DNA      linear      PAT 29-SEP-1999
DEFINITION Sequence 1 from patent US 5866684.
ACCESSION  AR031992
VERSION     AR031992.1 GI:5946281
KEYWORDS   Unknown.
SOURCE     Unknown.
ORGANISM   Unclassified.
REFERENCE  1 (bases 1 to 7475)
AUTHORS    Attwood,M.Richard., Hurst,D.Nigel., Jones,P.Stephen.,
            Kay,P.Brittain., Raynham,T.Michael. and Wilson,F.Xavier.
TITLE      Peptidyl inhibitors of viral proteases
JOURNAL    Patent: US 5866684-A 1 02-FEB-1999;
FEATURES   Location/Qualifiers
  source   1..7475
  source   /organism="unknown"
BASE COUNT 1853 a 1916 c 2039 g 1667 t
ORIGIN
Alignment Scores:      3.53      Length:      7475
Pred. No.:            49.00      Matches:      11
Score:                100.00%     Conservative: 0
Percent Similarity:    100.00%     Mismatches:  0
Best Local Similarity: 100.00%     Indels:      0
Query Match:          100.00%     Gaps:        0
DB:
US-09-965-594-26 (1-11) x AR031992 (1-7475)

QY      1 GlySerValValIleValGlyArgIleValLeu 11
Db      3451 GGCAGCGTGTTCATCTGTGGCAGGATCGTCTTG 3483

RESULT 5
LOCUS      AR207294              7475 bp      DNA      linear      PAT 20-JUN-2002
DEFINITION Sequence 1 from patent US 6372883.
ACCESSION  AR207294
VERSION     AR207294.1 GI:21506162
KEYWORDS   Unknown.
SOURCE     Unknown.
ORGANISM   Unclassified.
REFERENCE  1 (bases 1 to 7475)
AUTHORS    Attwood,M.Richard., Hurst,D.Nigel., Jones,P.Stephen.,
            Kay,P.Brittain., Raynham,T.Michael. and Wilson,F.Xavier.
TITLE      Antiviral medicaments
JOURNAL    Patent: US 6372883-A 1 16-APR-2002;

```

```

FEATURES             Location/Qualifiers
  source             1..7475
                    /organism="unknown"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:32644"
BASE COUNT          1853 a 1916 c 2039 g 1667 t
ORIGIN
Alignment Scores:      3.53      Length:      7475
Pred. No.:            49.00      Matches:      11
Score:                100.00%     Conservative: 0
Percent Similarity:    100.00%     Mismatches:  0
Best Local Similarity: 100.00%     Indels:      0
Query Match:          100.00%     Gaps:        0
DB:
US-09-965-594-26 (1-11) x AR207294 (1-7475)

QY      1 GlySerValValIleValGlyArgIleValLeu 11
Db      3451 GGCAGCGTGTTCATCTGTGGCAGGATCGTCTTG 3483

RESULT 6
LOCUS      HCJ238800              9033 bp      RNA      linear      VRL 18-AUG-1999
DEFINITION Hepatitis C virus type 1b complete genome, isolate NCI.
ACCESSION  AJ238800
VERSION     AJ238800.1 GI:5748510
KEYWORDS   complete genome; core protein; glycoprotein E1; glycoprotein E2;
            NS2 proteinase; NS3 proteinase/helicase; NS3/4A proteinase
            cofactor; NS4b protein; NS5A phosphoprotein; NS5B RNA dependant RNA
            polymerase; p7 peptide; polyprotein.
SOURCE     Hepatitis C virus type 1b
ORGANISM   Hepatitis C virus type 1b
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepacivirus.
REFERENCE  1
AUTHORS    Koch,J.O. and Bartenschlager,R.
TITLE      Modulation of hepatitis C virus NS5A hyperphosphorylation by
            nonstructural proteins NS3, NS4A, and NS4B
JOURNAL    J. Virol. 73 (9), 7138-7146 (1999)
MEDLINE    99370154
PUBMED     10438800
REFERENCE  2 (bases 1 to 9033)
AUTHORS    Bartenschlager,R.
TITLE      Direct Submission
JOURNAL
FEATURES   Location/Qualifiers
  source   1..9033
  source   /organism="Hepatitis C virus type 1b"
  source   /viroion
  source   /mol_type="genomic RNA"
  source   /strain="RB"
  source   /isolate="NCI"
  source   /db_xref="taxon:31647"
  source   /country="Germany"
  source   1..9033
  source   /codon_start=1
  source   /product="polyprotein"
  source   /protein_id="CA53095.1"
  source   /db_xref="GI:5748511"
  source   /translation="MSTNPKPQKTKRNTNRRPDQVKFPGGGQIVGGVYLLPRRGPRL
            GYVTRKTSERQPRROPQIPKARQEGRAWAQGPWPPLYGNEGGLGWAGWLLSPRG
            SRPSMGTPDRRSRLNGVIDTLCGFADLMGLPLVCGAPLGARALAHGVRVLED
            GYNATGNLPGCSFSIFLLALLSCLTIPASAYEVNRVSGVHYVNDNSNASIVYEAD
            NIMHTPGVCYPCYRENNSSRCWALPTLAARNASVPTTIRHVDLLVGAAALCSAMY
            VQDLGCVFLVAQLFTFSPRRHTVODCNCSYIPGHVTHGRMAWMMNMWSTPAALVW
            SGLLRIPQAVDVAAGHWGLAGLAYISGVGNWAKVLIVMLIFAGVDGTVTGTM
            AKNTIGTILSPGSSQKIQLVNTNGSHINRTALNCNDNLNTGLFALALFYVVKFNSS
            GCPERMASGSPIDAFQAGGPTTYNESHSSDORPYCWHYAPRPGIVPAOVCVPYC
            FTSPVVTGTTDRFGVPTYSWGENETDVLNNTRRPQGNWFGCTWMSNGTGTCKGG
            PCNIGGSKLTCTPDCFRKHEATYTKCGSGPWLIPRCLVHPYPRLMWHYFCTVTF
            TIFKVMYGGVEHRLAEACNMTGRERCNLEDRSRSELSPLLLSTTEMOVLPCSETTL

```


CDS

311..9343

```
/codon_start=1
/product="polyprotein"
/protein_id="BAB18805.1"
/db_xref="GI:11559451"
/translation="MSTNPKPQKTKRNTNRPPQDVKFPGGGQIVGGVYLLPRRGPRL
GVRATKSPRQPRGQIPKAPQEGRAWAGPYPLVGPALGGAGLWGLSPRG
SRPSNGPTDPRRSNLKQVDTLTCGFADLMGYPLVGPALGGAGLWGLSPRG
GVNATGNLPGCSFSLFALLSCLTIPVSAYEVNYSVDYVYVNDCSNYSIYEAD
VLMHTPGVPCVQRENSRSCVALPTLAAANSIPITIIIRHVDLLVGAACASMT
VDLCGSVFLVSQTSFPRRYETVQDCNSIYFHVSGHRMAMDMNMSPTALVY
SOLLIPQAVMDVMVTAHMGVLAGLAYSIMYNAKVLIVMLLFGVDTGTYTSGAAS
GRTMSFASLSSGSSQNIQLINTGSHINRTALNCNESLNTGFLAAGTYTSGAAS
GCPERMASCRPTDEAGWGPITTHATPVSDQPCWHYAPQCGIYPALQVCGPYVC
FTPSVVGTTDRDSAGPYNGENETDVLINNTRPPOGWNFGCTWNSTGCTKCGG
PCNIGVGNNTLTCTDPCF8KHPATYAKGSGFWLTPRCWVDPYPLRHYCTVNF
TIFKVMYGVGVEHRLNAACNWRGDCNREDDRSLSPLLSLITTEWQVLPCTFTL
PALSTGLHLHONIVDVOYLGVGSVAVSVVIRWEYVLLLLFLLADWAGACACMLL
IAQAPALNVLVLAAGLAGHILSVFCCAATYKGLAFGAAYFYVYVLPALL
LALLAPRAYAMDREMAACGGVFGVGLALLTSPHYKVFGLKLMQLQITRAEHL
IQWLPPLNVRGDRDAILLTCAVHPELIFEITKILLAI FGLMVLQAGLRVYFVR
AQLVRAChLVKVSQGOYVOMLMLAALTGTYYNHLTPLRDMAHAGLRDLAVAE
PVLSDMETKILITWADTAAGCDLLAGLVPVSARREILLGPADGFCQWRLLAPIT
AYSQTRGLLGCITSLIGRDNKQNGEVQVYSTATQSFLATCVNGVCWLVYHAGTK
TLAGKPGFTIQTHTYDQDLVGMQAPGARMTPTCGSSDLYLVTRHADYIPVRRRG
DSRGSLLSPRISYLGSGGGLPLCGHVVGIFRAAVCTRGVAKAVDFVPVSMET
MRSPVFTDSSPAPVQFQVHLHAPTSGSKSTKPAAYAAQGYKVLVNPVSAATL
CFGAYMSKHGVDPNIRGTGVTITTGATITVSTYKFLADGCGGAYDIIIMCECHS
TDSYSLIGTVDQAEATAGARLVATATPPGVTVPHPNIEVALSNTGEIIFYCK
AIPVIEYKGGHILFCHSKKCKDELATKLSALGINAVAYVGLDYSVITPSGVVVA
TUALMTGTGDFSDVIDNTCVTQVDFSLDPTTETITMPQDASRSQRRGTGRG
RGTIRFTVTPGSDSSVLCEDYDAGCANYELTPAETSVRRLRLNTPLGPFVQ
DLEFWESVFTGLTHIDAFUSOTKQAGENLPYVAYOATVCAQAAPPPSMDMMKC
LVRLKPTQGTPTLLYRLGAVONEVLTHTPIKYITMCMAADLEVVTSTVYLVGGVLA
ALAAVLTGTVVIVIRVILSGRPVAVDPREVLYOEDEMEECAASHLYPIEONGHAE
QFQKALIGLITATQAAEAAPVYESKWRLEGFWAKHWNFIISIDYLAGLSLPGN
PAISLMAFTSITSLPTONTLMENILGCHVAQLAPPSSASAPVAGIAGAAGVSI
GLGKVLVDILAGVAGALVAFKVMSEMPSTEDVNLPLPAILSPGALVGVVCAA
ILURHVGEGAGVQMMNRLIFASNGHVSPTHYVPESDAAARVQIISLTIITQLLK
RLHQWINECSTPCSGWLDVMDICAVLDFKTLQSKLLPRLPGVPPFSCQRYK
GVWRGDVHHTTCCGAGISGHVKNMSRIVGPKTCSNWTGCTFPINAYTIGCTPSP
APNYSKALMRVAEYEVTVRQFVHYVTLGTLTONVPCPOVPAFEFTLDGVRLHR
YAPACKPLLRDEVTQVGLNOYPVGSOLPCEPEPDVTLTSMLTOPHITAEARRL
ARGSPSLASSANTDGLSKAKCTCHNAPDADLEALLMQEKGNGNITRVESN
KVYILDSPEPLRAEDGREVSPAEILARTKFPPLAPINWADINPLLESSEADST
VPPVHVGCLPTKAPPVPPRRKRVVLTETVSSALAEATKAFKSSGVSSAVDST
ATAPDQDLDDGSDGVSSEMPPEGEFDPDLSGWSVTSVEEGEDVCCSMS
YTWTGALITPCAAEESKLPINALSNLHNLVYATTSASQOROKKRVTFDRQLVD
HYRDLKEMAKASTVKARLLSVEACKLTPHRSARKFGYGAQDKVRSKKAHLHI
HSVWDLLEDTEPTIOTTIMAKNEVFCVQEKGRKPARLIVFPDLGVRYCEKALYD
VYSTLPQAVMGSSYGFQYSPGORVEFLVNAKSKNPMGFAYDTRCESVTEKDIY
EESIYQCDLAPEARLAIIRSLTERLYVGGPHINSKQNGCYRRCRASGLVITSCGNTL
TCYLKASACRAKURDKTMVCGDOLVVICESAGTQEDAAASLRYFTTEAMTRYSAAPG
DPQPEYDELELITSSVNVAHADANKRVYLTDRDPTPLSRAAETARHTPYNSWL
GNIMYAPTLMWMTLTHFFSILIAQOELEKALDCQIYGAISIEPLDLPQIQRHL
GLASFSLATYSGEINRVASCLRGVPLVRWRHARSVRKLSLOGGGAAGICGKYL
FNWAVRTKLKLTPIPAASRLDLSWVAGYSGGDIYHSVRSRPRFWLMLCLLLSVCV
GYLLPNR"
```

BASE COUNT 1889 a 2842 c 2655 g 1958 t

ORIGIN

Alignment Scores:
Pred. No.: 4,49 Length: 9344
Score: 49,00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-09-965-594-26 (1-11) x AB049092 (1-9344)

Qy

1 GlySerValIleValGlyArgIleValLeu 11

|||||

Db 5342 GCACGGTGGTCATTCTGGCAGGATCGTCTG 5374

RESULT 8

HPCCGENOM

LOCUS

DEFINITION

Hepatitis C virus complete genome.

VERSION

102836

KEYWORDS

complete genome; viral genome.

SOURCE

Hepatitis C virus

ORGANISM

Hepatitis C virus

REFERENCE

1 (bases 1 to 9400)

AUTHORS

Bi.S.-l., Bai.X.-H., Margolis.H.S. and Liu.C.-B.

TITLE

Complete hepatitis C virus genome cDNA isolated from mainland of China

JOURNAL

Unpublished (1992)

COMMENT

Original source text: Hepatitis C virus (individual isolate China, strain HeBei) (library: NC1-18) cDNA to genomic RNA.

FEATURES

Location/Qualifiers

source

1..9400

/organism="Hepatitis C virus"

/mol_type="genomic RNA"

/strain="HeBei"

/isolate="China"

/db_xref="taxon:11103"

/tissue.lib="NC1-18"

355..9384

mRNA

1920 a 2796 c 2650 g 2034 t

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.: 4,52 Length: 9400

Score: 49,00 Matches: 11

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 14 Gaps: 0

US-09-965-594-26 (1-11) x HPCCGENOM (1-9400)

Qy

1 GlySerValIleValGlyArgIleValLeu 11

|||||

Db 5362 GCACGGTGGTCATTCTGGTGGATTTGCTTG 5394

RESULT 9

HPCRNA

LOCUS

DEFINITION

Hepatitis C virus RNA, complete genome sequence.

ACCESSION

D10934

VERSION

D10934.1 GI:471116

KEYWORDS

C; E; NS1/E2; NS2; NS3; NS4; NS5; polyprotein precursor.

SOURCE

Hepatitis C virus

ORGANISM

Hepatitis C virus

REFERENCE

1 (bases 1 to 9456)

AUTHORS

Wang.Y., Okamoto.H., Tsuda.F., Nagayama.R., Tao.Q.M. and Mishiro.S.

TITLE

Prevalence, genotypes, and an isolate (HC-C2) of hepatitis C virus in Chinese patients with liver disease

J. Med. Virol. 40 (3), 254-260 (1993)

JOURNAL

Immunology Division

Jichi Medical School

Kawachi-gun

Tochigi 329-04

Japan

Phone: 0285-44-2111 x3334

Fax: 0285-44-1557.

CDS

311..9343

```
/codon_start=1
/product="polyprotein"
/protein_id="BAB18805.1"
/db_xref="GI:11559451"
/translation="MSTNPKPQKTKRNTNRPPQDVKFPGGGQIVGGVYLLPRRGPRL
GVRATKSPRQPRGQIPKAPQEGRAWAGPYPLVGPALGGAGLWGLSPRG
SRPSNGPTDPRRSNLKQVDTLTCGFADLMGYPLVGPALGGAGLWGLSPRG
GVNATGNLPGCSFSLFALLSCLTIPVSAYEVNYSVDYVYVNDCSNYSIYEAD
VLMHTPGVPCVQRENSRSCVALPTLAAANSIPITIIIRHVDLLVGAACASMT
VDLCGSVFLVSQTSFPRRYETVQDCNSIYFHVSGHRMAMDMNMSPTALVY
SOLLIPQAVMDVMVTAHMGVLAGLAYSIMYNAKVLIVMLLFGVDTGTYTSGAAS
GRTMSFASLSSGSSQNIQLINTGSHINRTALNCNESLNTGFLAAGTYTSGAAS
GCPERMASCRPTDEAGWGPITTHATPVSDQPCWHYAPQCGIYPALQVCGPYVC
FTPSVVGTTDRDSAGPYNGENETDVLINNTRPPOGWNFGCTWNSTGCTKCGG
PCNIGVGNNTLTCTDPCF8KHPATYAKGSGFWLTPRCWVDPYPLRHYCTVNF
TIFKVMYGVGVEHRLNAACNWRGDCNREDDRSLSPLLSLITTEWQVLPCTFTL
PALSTGLHLHONIVDVOYLGVGSVAVSVVIRWEYVLLLLFLLADWAGACACMLL
IAQAPALNVLVLAAGLAGHILSVFCCAATYKGLAFGAAYFYVYVLPALL
LALLAPRAYAMDREMAACGGVFGVGLALLTSPHYKVFGLKLMQLQITRAEHL
IQWLPPLNVRGDRDAILLTCAVHPELIFEITKILLAI FGLMVLQAGLRVYFVR
AQLVRAChLVKVSQGOYVOMLMLAALTGTYYNHLTPLRDMAHAGLRDLAVAE
PVLSDMETKILITWADTAAGCDLLAGLVPVSARREILLGPADGFCQWRLLAPIT
AYSQTRGLLGCITSLIGRDNKQNGEVQVYSTATQSFLATCVNGVCWLVYHAGTK
TLAGKPGFTIQTHTYDQDLVGMQAPGARMTPTCGSSDLYLVTRHADYIPVRRRG
DSRGSLLSPRISYLGSGGGLPLCGHVVGIFRAAVCTRGVAKAVDFVPVSMET
MRSPVFTDSSPAPVQFQVHLHAPTSGSKSTKPAAYAAQGYKVLVNPVSAATL
CFGAYMSKHGVDPNIRGTGVTITTGATITVSTYKFLADGCGGAYDIIIMCECHS
TDSYSLIGTVDQAEATAGARLVATATPPGVTVPHPNIEVALSNTGEIIFYCK
AIPVIEYKGGHILFCHSKKCKDELATKLSALGINAVAYVGLDYSVITPSGVVVA
TUALMTGTGDFSDVIDNTCVTQVDFSLDPTTETITMPQDASRSQRRGTGRG
RGTIRFTVTPGSDSSVLCEDYDAGCANYELTPAETSVRRLRLNTPLGPFVQ
DLEFWESVFTGLTHIDAFUSOTKQAGENLPYVAYOATVCAQAAPPPSMDMMKC
LVRLKPTQGTPTLLYRLGAVONEVLTHTPIKYITMCMAADLEVVTSTVYLVGGVLA
ALAAVLTGTVVIVIRVILSGRPVAVDPREVLYOEDEMEECAASHLYPIEONGHAE
QFQKALIGLITATQAAEAAPVYESKWRLEGFWAKHWNFIISIDYLAGLSLPGN
PAISLMAFTSITSLPTONTLMENILGCHVAQLAPPSSASAPVAGIAGAAGVSI
GLGKVLVDILAGVAGALVAFKVMSEMPSTEDVNLPLPAILSPGALVGVVCAA
ILURHVGEGAGVQMMNRLIFASNGHVSPTHYVPESDAAARVQIISLTIITQLLK
RLHQWINECSTPCSGWLDVMDICAVLDFKTLQSKLLPRLPGVPPFSCQRYK
GVWRGDVHHTTCCGAGISGHVKNMSRIVGPKTCSNWTGCTFPINAYTIGCTPSP
APNYSKALMRVAEYEVTVRQFVHYVTLGTLTONVPCPOVPAFEFTLDGVRLHR
YAPACKPLLRDEVTQVGLNOYPVGSOLPCEPEPDVTLTSMLTOPHITAEARRL
ARGSPSLASSANTDGLSKAKCTCHNAPDADLEALLMQEKGNGNITRVESN
KVYILDSPEPLRAEDGREVSPAEILARTKFPPLAPINWADINPLLESSEADST
VPPVHVGCLPTKAPPVPPRRKRVVLTETVSSALAEATKAFKSSGVSSAVDST
ATAPDQDLDDGSDGVSSEMPPEGEFDPDLSGWSVTSVEEGEDVCCSMS
YTWTGALITPCAAEESKLPINALSNLHNLVYATTSASQOROKKRVTFDRQLVD
HYRDLKEMAKASTVKARLLSVEACKLTPHRSARKFGYGAQDKVRSKKAHLHI
HSVWDLLEDTEPTIOTTIMAKNEVFCVQEKGRKPARLIVFPDLGVRYCEKALYD
VYSTLPQAVMGSSYGFQYSPGORVEFLVNAKSKNPMGFAYDTRCESVTEKDIY
EESIYQCDLAPEARLAIIRSLTERLYVGGPHINSKQNGCYRRCRASGLVITSCGNTL
TCYLKASACRAKURDKTMVCGDOLVVICESAGTQEDAAASLRYFTTEAMTRYSAAPG
DPQPEYDELELITSSVNVAHADANKRVYLTDRDPTPLSRAAETARHTPYNSWL
GNIMYAPTLMWMTLTHFFSILIAQOELEKALDCQIYGAISIEPLDLPQIQRHL
GLASFSLATYSGEINRVASCLRGVPLVRWRHARSVRKLSLOGGGAAGICGKYL
FNWAVRTKLKLTPIPAASRLDLSWVAGYSGGDIYHSVRSRPRFWLMLCLLLSVCV
GYLLPNR"
```

BASE COUNT 1889 a 2842 c 2655 g 1958 t

ORIGIN

Alignment Scores:
Pred. No.: 4,49 Length: 9344
Score: 49,00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-09-965-594-26 (1-11) x AB049092 (1-9344)

Qy

1 GlySerValIleValGlyArgIleValLeu 11

|||||

Score:	49.00	Matches:	11
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	14	Gaps:	0
US-09-965-594-26 (1-11) x HPCRNA (1-9456)			
QY	1	GlySerValValIleValGlyArgIleValLeu	11
DB	5373	GGCAGTGGTCATTTGGCGAGGATCGTCTTG	5405
RESULT 10			
LOCUS	AF176573	9600 bp	RNA
DEFINITION	Hepatitis C virus polyprotein precursor, gene, complete cds.		
ACCESSION	AF176573		
VERSION	AF176573.1	GI:5738246	
KEYWORDS	Hepatitis C virus		
SOURCE	Hepatitis C virus		
ORGANISM	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.		
REFERENCE	1 (bases 1 to 9600)		
AUTHORS	Mokhonov, V.V., Samokhvalov, E.I., Novikov, D.V., Shatalov, A.G. and Prilipov, A.G.		
TITLE	Molecular cloning HCV Russian isolate lb from the serum of patient with acute hepatitis		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 9600)		
AUTHORS	Mokhonov, V.V., Samokhvalov, E.I., Novikov, D.V., Shatalov, A.G. and Prilipov, A.G.		
TITLE	Direct Submission		
JOURNAL	Submitted (09-AUG-1999) Molecular Genetics, Ivanovsky Virology Institute, Gamaleya Str., 16, Moscow 123098, Russia		
FEATURES	Location/Qualifiers		
source	1..9600		
	/organism="Hepatitis C virus"		
	/mol_type="genomic RNA"		
	/strain="274933R"		
	/isolate="1b"		
	/db_xref="taxon:11103"		
	/country="Russia"		
	/note="Isolated from acute hepatitis patient serum"		
5'UTR	1..341		
CDS	342..9374		
	/note="cleaved into C, E1, E2, p7, NS2, NS3, NS4a, NS4b, NS5a, and NS5b proteins"		
	/codon_start=1		
	/product="polyprotein precursor"		
	/protein_id="A4D50312.1"		
	/db_xref="GI:5738247"		
	/translation="MSTNPKPQKTKRNTNRPPQDVKFPGGQIVGGVYLLPRGPRLGYRATKTSERSQRRQPIPKARREGRRAQPGYPWPLYNKGWAGHLLSPRGSRPSMGNDPRRRNRNLGKVIDTLTCGFADLMGYILVPGAPLGGAALAHGVRAEDGVNATGLPGCSFIFLLALLSCLTIPASAEVYNASGVYHVNDCSNISVYEADVIMHNGPCVPCRENNRSQVALTIPAAARNASVPTTIRRHVDLLVGAALCSAMYGMHNGCVFLSQSLTFSPRHETVQDCNSIYGHITGHRMAMDMNNMSPITALVYQALLRIPAQMDVAGAHVGLAGLAYSVMGNMAKVLIVLLFAGVDGNTVYTGGAARGASGITSLFSGRPSQKIOLVNTNGSHINRTALNCNDSFNTGLAALAFYARENSSGCPERMASCRSIDFDQGWGPIYVQDSDQRPYCHYPPRCGIVPASEVCGPYVCPTPSPVVTGRLGQPYTNNGENETVLLNNTPRPQGNWFGCTNNITTFGTNKTGGFPNCNTAGGAGNLTCTPTDFPKHPEATYKCGSPMLTPCLVDYPLRYHPTCTVNFPTFYKMTYGVGVEHRLDRAQNMWTCERCALEDROSELSPLLSLSTEWQLPSCFTTLPALSTGLHLHRTIVDQVLYIGISVAFKWEYVLLFLLIADARVACCLMMLLIAQAELANLNVNAAVAGVGHLSFLVFECAATYIKGRVPGAAFYGVWPLILLLALPPRAYAMARENASCGGAVFGLALLTISPYKVFRLRIWLVYFIRAEHLQVWLPPLNVYGRDAIILLCAAHPELIFDITKLLAIIRGPTNVIAQAGTKPKYFVRAGGLRGLVPLVAGVGHVQMAFMKLAALTGYVYDHLTPLRDWAHTGURDLAVAYEVFSDMETKTIITWGAATACGDIILGLPSARREKILGLPADSLEGQWRLLAPITAYSGQTLTGCGIITSLTRDKNOVEGVSVYSTATQVLCATVNGCYMTVYHGASGLTLAGPKGTITOMYTNVDQVLGWAPSGARSLTPTCGSSDLVLTARYADIVYRARGDSRGLSPRPVYLLSGSGGPLLPCSGHAGIFPRAAVCTRGVAKAVYVLPVPSMETTMRSPVTDNSPPAVPOTFOVAHLHAPHTGSGSKTKVPAAYAGQYKVLVLPNSVAATLCEGAYMSAHGIDPNTGTGVTITGAPITISTYKELADGGCGSGAGLIIDCECHSDTSTIIIGLHVLQDAGTAGRLVLAITAPPGSVYVPHSNIEVALYSTIGELPFTYKALPITIIGKGRHLIFCHSKKCKDELAALKSALGNAYVYRGIDVSIPTSGDVVYVATDALMTGTFDQSVIDCNDTPTVDPSLDTPTIETTVPDQVNSRRRGRTGRRGIRYFTYPTGERSGMDSVLCECYDAGCAEYELTPAETSVRLRAYLNTPLPVQGLHLEWESVFTGLHIDAHFLSQKQAGDNFPLVAYATVCARQAAPPSPWQMKKCLRLPRLTGHPTLLYRLGAVQNEVLTHTPTKYIMACMSADLEVTSTWLVGGVLAALAYCLLTGTSVIVGRVLSGRNAVIPDREVLQERDEMEGSHLPYIEQGMOLAEAFYQKALGLQIATQAAAPAVVESKMRALETFWAKHMWNTISGVYILAGLSLTPGNPAIASLAFTASVTSPLTQSLTNFLNLGGWVAQALAPPASAFYAGIAGAAVGSIGKGLVYDILAGYAGVAGALVAFKMSGETPSAEDVNLPLPAILSPGALVGVCAAIGRRVHGEGEAVOMNRLIAFASRGNHVSPTHYVPESDAAARTQLSLLITQLIKRUHONEDCSTPCGSGWLRIDMWICSVLTDKTLQSKLLPRLPGVPFPFGCGYKGVWRGDMOTTCPCGAQITGHWKNGSMRIYGPKTCSNTWHGTFPINAYITGCTPSPAPNYSRALRWRAEYEVTVGVDPHYITGNTONIKPCQVPAPEFTFVDGVRLHRYACKPVLRVEVDVQGLNOYPVGSQLPDPEPDVAVLSMLDPSHITAEAKRRLYAGSPSLASSASLAPSALUKATCTHDDSPDADLLEANILMQEMGNITRVESNKKVIDLSEPLRAEEDEREVSAAEILURKTRFFPAMPVAPARYNPILLESWKDQYVPVYHGGCLPPTPKAPITPPRRKRTVLTSTVSSALAEATITFGSSGSVADSTYATAPDQTSNDRSDRESAESYSSWPLEGEPDGLSDGMSVTSSEASGVVCCSSATWMTALITPCAARESPLINPLSNLHRHNMVYATTSRSASLROKKVTFDRLQVIDDYRDLVEMKAKASTVAKLISLEACKLTPHSKSKFCYGAADKVRNLSKAVNHLRYWDLLEDVETPIDTITMAKNEVFCVQPEKGRKPARLIVFDLVYVCEKMALYDVYSTLPADMGSSYGFYQGVORFVLVWKKSKPMGFSYDFTRCFDSTVETENDIRIESIYOCDDLAPEAKQAKLSUTERYIGGLTNSKNGCGYRRCRASGVLTFTSCGNTCYLKASACAAKQAQCTLVNGDGLVVICESAGTODAAALRVFTTEAMRTYASPGCDPQPEYDELIETSCGSNVYSAHADSKRRVYLTDRPTPLARAADEARTHTPVNSWLGNIINYPATLARMILMTWTFHLSILIAQOELEKALECOIYACYSIEPLDLPLIERHLGLSFLSHSYSGEINRVASCLRKLGVPLVYWRHRSRVAKLLSOGGGAATCGKYLFWAVRTKLKLTPIPAASRLDLSGWVAGYGGGDIYLSLSRARPRWFMLCLLLLSGVVGYLLPNR"		
3'UTR	9375..9456		
ORIGIN	BASE COUNT 1909 a 2811 c 2668 g 2068 t		
Alignment Scores:	4.55		
Pred. No.:	9456		
Length:	9456		
with a T-stretch of 45 bp."			

MRSPVFTDNSSPPAVPQTFQVAHLHAPTGSCKSTKVPAAYAAQGYKVLVNPVAAATL,
GFGVMSAHGVDNQLRTGVRTITTTGAPITYSTYCKFLADGGCGGAYDIICDECHS
TDSITLIGTVGLVQDLNAGARLVATATPPGVSVPVPHNIEGIALSNTGIEPFYCK
AIPBILKGGHBLIFCHSKKCDLAKLSGLNNAVAYRGLDVSVPISGNNVVA
TDALMTGTFDNCVDTCVDFSLDPTFTIETITVPODAVSQRQRGTGRG
RRGIYRVTPGERSPGSDSVLCEDDAGCAMELTPAETSVRLRAYLNTPLGPVCO
DHLFEWSVFTGLTHIDAHFSLQTKOAGDNFYPYLAOATVCARAQAPFSWQMMKC
LIRLPTLHGPTPLLYLGAQVNETTTLHPITKYIMACMSADLEVSTFWLVGGVLA
ALAAVCLTGSVIVRGLVSLGKPAIPDREVLYOEFDEMECAASHVLYIEOGMOLAE
OFKALGLLOTATKQAEAPAVVESKWRALFAKHMWFISSVOYLAGLSLPGN
PAIASLMAFTASISPLTQYTLLENILGGWVAQAQLAPPSSAASAFVAGAGIAGAAGSI
GLKVLVDILAGVAGAGALVAFKMGSEMPSTEDLVNLLPALILSPCALVGVVCAA
ILRRHVGEAGAGOMMNLIAFASGNHVSPTHYVPESDAAARVQILSNUTIQOLLK
RLHWMINDCSTPCSGSLRDVMDWICVLTDFKTLWRSKLLPRLPGVPFLSCQGRK
GWMRGDMOTTCPCGAGIAGHNGSMRIVGPRCTSNTHGTFPIINAYTGPCTPSP
APNYSRALWRVAEEYVEITRVGDFHYVTGMTDNVCKPCOVPAPEFTEVDGVLHR
YAPVCKPILHEDVTQVGLNOYLGSQLPCPEPEPDVAVLTSLDPSHIFTAKRRL
KNGSPSLASSASQSLKATCTTTRHSDPADLIEANLLWQEMGNITRVESEN
KVWILDSFDPLRAEEDEREVPAPILKTRKFSAMPIMAWPDYNPLLESWKPDDY
VPPVHHGGLPPTKYVPIPPRRKRTVTLVSTVSSALAEATKTFGSESSAYDSGT
AAASDPQSDNGTSDSVESYSSMPLEGEFDPDLSGMSVSEASESSVDCSWS
YTWTCALLTPCAAESQLPINALSNLRLHNLVYATTSRASOROKKVTEDRLQVLD
HYOVVLKEMKAKSTVKALLSVFEACKLTPHSAKSKFCGYGKDVNLSKXAVNH
HVSXKDLLEDETPTIDTIMAKNEVFCQPEKGGKPARLIVFDLGVRCERKALVD
EESIYQCCDLAPARQAIRSLTERLYGGPLTNSKQNGCYRRCAAGVLTSCGNL
TCYLKATAACRAAKRQDCTLLVGDOLVICESAGTQBDAAASLRFTEATRYAPGP
DLPOPEYDELEITSRSSVVAHADSGRVTYLTDRPTPLARAWEAARHTPNSWL
GNIMYAPTEILWARMILMTFFSILLFQBLEKALDCQIYGAYSTIEPLDLPOIIFORLH
GLSAFLSHSYPGENRNASCLKGLVPLRAWRHRARNVRKILLSGGGRATCGKYL
FNMAVTRKLKLTPIPAASQDLSDGFWAGYGGGDIYHLSLRARPWPMCLLILFVG
GLYLPNR"

9375..9600

3'UTR BASE COUNT 1917 a 2887 c 2696 g 2100 t
ORIGIN

Alignment Scores:
Pred. No.: 4.62 Length: 9600
Score: 49.00 Matches: 11
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-09-965-594-26 (1-11) x AF176573 (1-9600)

Qy 1 GlySerValValIleValGlyArgIleValLeu 11
|||||
Db 5373 GCGAGCGTGCATCTGTCGGCAGGATCGTCGTTG 5405

RESULT 11
AR179057
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

1 (bases 1 to 12734)
Sequence
1 from patent US 6326137.
AR179057
AR179057.1 GI:20220612

Unknown.
Unknown.
Unclassified.
REFERENCE
1 (bases 1 to 12734)
Hepatitis C virus protease-dependent chimeric pestivirus
Patent: US 6326137-A 1 04-DEC-2001;

Location/Qualifiers
1..12734
/organism="unknown"
4032 a 2604 c 3295 g 2803 t

BASE COUNT
ORIGIN

Alignment Scores:
Pred. No.: 6.27 Length: 12734
Score: 49.00 Matches: 11

Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-965-594-26 (1-11) x AR179057 (1-12734)

Qy 1 GlySerValValIleValGlyArgIleValLeu 11
|||||
Db 413 GGTAGTGTCTTATTGTTGGTAGAATTGTTTA 445

RESULT 12
AF268278
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

1 (bases 1 to 12734)
Pestivirus type 1, complete genome.
AF268278
AF268278.1 GI:9049956

Pestivirus type 1
Pestivirus type 1
Viruses: ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Pestivirus.

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED

1 (bases 1 to 12734)
Lai, V.C., Zhong, W., Skelton, A., Ingravallo, P., Vassilev, V.,
Donis, R.O., Hong, Z., and Lau, J.Y.
Generation and characterization of a hepatitis C virus NS3
protease-dependent bovine viral diarrhoea virus
J. Virol. 74 (14), 6339-6347 (2000)
20323484
10864644

2 (bases 1 to 12734)
Lai, V.C.H. and Hong, Z.
Direct Submission
Submitted (16-MAY-2000) Antiviral Therapy, Schering-Plough Research
Institute, 2015 Galloping Hill Road, Kenilworth, NJ 07033-0539, USA

Location/Qualifiers
1..12734
/organism="Pestivirus type 1"
/mol_type="genomic RNA"
/db_xref="taxon:11099"

5'UTR
CDS
386..12508
/codon_start=1
/product="polyprotein"
/protein_id="AAF82566.1"
/db_xref="GI:9049957"

/translation="MELNTNEGSGSVIVGRIVLSSGSGSITACAAQOTRGLGCKITSL
TGRDKNQGEVQIVSTATQTFLATCINGVCTVYHGAGTRTITASPRGPVIOYNTVD
QDLGWPAPOGSRSLTPTCTGSDLYLVTRHANYIPVRRGRSGSLLSPRLSYLKG
SSGGLLCPAGHAGLFPRAVCTRGVAKAVDFIPVENLETTTRSGSAGATFEDVCCSM
SYSDTKEGATKKTATQPKDLERGMKIVPESEKSKTKPPDATTIVESGKTVQVRKK
GKTSKRTQDLGHNNKKNKPKESKKLEKALLAIIAIVLFQVTMGENTITQWNLQNG
TEGIQAMFGVNRSLHGWPEKICTGVFSLHATIDIELTTHGMDSASEKTYTCGR
LQREHMKGCMNYPNIEPLVNMRTQANLQDPPRECACTVCRDASDLNAVTTQA
RDSPTPLTGCKKGNFAGILMRGPCNFAIASDVLFEKHERISMFOFTLLVLDGL
TNSLEGAROGTAKLTWLGKQGLGKLENKSKTWEGAVAYAPYCDVORPKIGYIAT
XNCTPACLPNTKIYVPGKFDNADCKILHEMGHLSVLLSLVLSVDFAPETASV
MYLHFSIPSHVDVMDCKTQNLTVLTADVIPGSVNNLGNKWCIRPNMPVET
TVLAFEEVSQVVKLRLALDRTRIWAATTTAFCLVLYRGQMVQOILMLLIT
GVQGLDCKPEFSYAIADDERIGQGAELTTTKEYSPGKLEDLTMVIAWEDGKLM
YLOCTRETRYLAIALTRALPTSVFKFLDGRKQEDVEMDNFEGCLPCDAKPIV
RGKENTLLNGPAFQWVCPITGWTCTVSTFNMDTLATTVRTVRSKPPHQCIT
QKNLGEDLHNCILGGNMTCPGQQLLYKGSIESCKWCGQPFRESGLPHYKGLKCL
ENETGYRLVDSSTCNREGVAIVFGDLKRLKIKGTIVQIANDTLKMGPCRPETELSS
EGPVETACTFNTKLNKIFEPDRSFQOYMLKGEYVQVANGNLLTHNNIEVYTFLL
LYLLREESYKVKVLLYHILVHPHKSIVILLMIGDVVKADSGOGEYLGKIDLCFT
TVLIVGLIILARBDPTIVPLVTIMARLTETHPQVDIAVAVMTILLNYSYVD
YFRYKWLQCLISLVSGVFLIRSLIYLGRIEMPEVTPNWRPLTLLLYLISITVTR
WKVDVAGLLQCVPIILLVTLVADFTLLILPTVELVLYKLTVTTRDIERSWLG
IDTRVDSIYDVDSGEGVYIFPSRKAQGNFSILLPLIKATLISCVSSKWLITYMSY
LTDPMYMHKRVTEESGGTNIISRLVAALIELNMSMEESKGLKFLYLLSRLRN
LIINKRVNETVASHTGYGPKIMTIIKASTLSKSRHCIICTVCEGRENKGTGC

PKCGRGRDITCTMSLADEPERRHYKRIIFIREGFMGSCROGKRRRREMFEDREPKSAR
YCAECNRLHPAEGEDFWAESSMLGKITITYFALMDKATYDITWAGCORGISPDTHRV
PCHISFSPRQVNFQVYQATGOLFRLNLPVATKVKMLMVGNGEIGNLHL
GHLSPACVKKITIEKCHINILDKLTAFGIMPRGTTPRAPVPFPTSLKVRGLE
TGWHTQGGISSDHYTAGKDLLVCDMGRITRVYCQNNRLTDETEYGVKDGCPD
GARCYPVPEANVSGSKGAVHLOKTGGEFTCVASGTPAFFDLKLNKLSGSLPFE
ASSGRVTRVVKNEESKPTKINSIGTQVSKNTADLTWKKITSMRNGDFKQITLA
TGAGRTTELPRKAVLIEIGRKRVLVLIPLRAAESVQVYMRKHPISFNLRIQDMKE
GMATGTVASGYKTCOMPORPKRAAMEVSYIFLDEYHCATPEOLAIIKIHRESE
IRVANTATPAGSVTTGOKHPIEEELAPEVMKGEDSGQLDIAGLIKIPDEKGNM
LVFVTRNMAVEAKLAKAGNSGYTSGEDPANLRVYVTSQSPVIVATNAIESGVT
LPDLOITDGLCKRVRVSSKIPFIVTGLKMAVTVGEQQRGRVGRVKGPRYIR
SOETATGDKYLDLOAQRYGIEDINVTYSKREMYDMSLEYEDSLLIQLEILNN
LILISDPAAVKINARTDPEPIQLANSEYEVQVFLPKIRNGEVDTDFYSEAFLL
ARKLGEDVPVYIATDEDELDVLLGDPDNGOQVETGKALQVGTLSAENALL
VALFGYVQALSKRHPMTIDITYIEDORLEDTTHLOYAPNAIKDCTETELKELAS
GVDEKIMGASIDYAGGLEFVKSQAEKIKIAPLFKENEAEAKGYQKQIDSLIENKEE
IIRYGLMHTHIALYKISLAARLGHETAFATLVKLWLAFGESVSDHQVAAVLDLVVYV
MKNPSPGDSUETQOGRFRVASFISALATYTYKWNHNSKVVPEPALAYPATSA
LMFTPTRESVYILSTIYKTYLSIRKGSGLGTGTSAAEILSONPVSVGISVM
LGVGALAHNAIETSEOKRTLLMKVFNKFLDOAATDELVENPEKIIIMAFEAQVOTI
GNPLRLIYHLYGVYKWEAKELSEKTAGNLTFLIMEAFELGMDSQKIRNLSGN
YILDLYGLHKOINRGLKMWLGMAPESCDMTPSDERILRPLTDNYLRVETRCPGY
EMKAKNVGGLTKVKEESGPFELCRNRPGRVNYRYKYDDNLEIRIPVAKLEGQVE
HYKGVTAIDYKSKMLLATDWEVHGVIITRIAKRYTGVDGFGNAGIIGDEPNHRLV
ERDCATITNTVOFLMKMGCAFTYDLTISNLTRELIELVHRNNEKEEIPATVYTWL
AYTFVNEYDGTIKPVLGERVPIQVVDINLOPEQVDTSEVGTIIGRETMTTGTVP
VLEKPEPDSQNSVKIGLDEGNTPGGCIQHTLTETEHNRDARPTFIIWLSRNSIS
NRAKTARNLTNGDNPREDIMAAAGMLVVALRDVDPSEMDYDFKGTFLDREALE
ALSLOPKPQVTKKAVRNLIEQKQVVEIPNMFASDDPVEVALKNDKYVLVGDVGE
VDDOAKALCATQTRIKEVGSRTYAKLSSFLQASNKQMSLTLPFELLRLRCPAT
KNGKMSAYLOAGNPELGGVHLGTIPARRVKIHPYEAYLKLKDFIEEEKKPR
VADTVIREHKNWILKIRFQENLTKMLNPGKLSQDLREGKRNINHOIGTIMSS
AGTRLEKPIVRAQDTKTFFHEAIRDKIDKSENRONPELHNKLLIEFTIAQTLKHT
YGEVWEOLEAGNKGAGLEKKNIGEVLDSEKHLVEOLVRLDKAGRKIKYETAI
PKNERDVSDDAQDLVVEKPRVIOVPEAKTRLAITKVMYMWKQPVIPVGEK
TPLENFDKVRKENSFPNAVSDTKANDQVTSKOLQIGELQRTYKEMHKFI
DITDHTIEVITADGEVITIRNGQSGOPDTSAGNSMLNVTMTWAFCESTGVPIK
SENRVARHVCDDGLFITEKGLKFKANKQIHEAGKPKQITEGEMKMYAVFELL
IFCSHTPPVPMWSNDSHAGRDYAVILSKMATLSDSGERGTYAYEKAVAFSFL
MYSNPLVRRIICLLVLSQOPEDPKSHATYIYKQDPIGAYKDVIGRNLSELKATCFEK
LANLNLSTLGTWTKHTSKRIIQDCAVJGKEGNWLVNADRLISKTKCHLYIPDKGF
TLQGHVEQLOLRTETPNVGVGTERYKLGPIVNLRLKILLMTANGVSS"
12509...12734

3'UTR 4030 a 2608 c 3293 g 2802 t 1 others
BASE COUNT
ORIGIN

Alignment Scores:
Pred. No.: 6-27 Length: 12734
Score: 49.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-09-965-594-26 (1-11) x AF268278 (1-12734)

Qy 1 GlySerValValIleValGlyArgIleValLeu 11
|||||
Db 413 GGTAGTGTGTTATTGTTGTAGAAATTATTTA 445

RESULT 13
AR145217
LOCUS AR145217
DEFINITION Sequence 48 from patent US 6211338.
ACCESSION AR145217
VERSION AR145217.1 GI:15107084
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 75)
AUTHORS Malcolim,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.

Qy 1 GlySerValValIleValGlyArgIleValLeu 11
|||||
Db 13 GGTCTGTGTTATTGTTGTAGAAATTATTTA 45

RESULT 15
AR145197
LOCUS AR145197
DEFINITION Sequence 26 from patent US 6211338.
ACCESSION AR145197
VERSION AR145197.1 GI:15107064
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 78)
AUTHORS Malcolim,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.
TITLE Single-chain recombinant complexes of hepatitis C virus NS3
protease and NS4A cofactor peptide
JOURNAL Patent: US 6211338-A 26 03-APR-2001;

TITLE Single-chain recombinant complexes of hepatitis C virus NS3
JOURNAL protease and NS4A cofactor peptide
Patent: US 6211338-A 48 03-APR-2001;
FEATURES Location/Qualifiers
source
BASE COUNT 16 a 14 c 15 g 30 t
ORIGIN
Alignment Scores:
Pred. No.: 0.0408 Length: 75
Score: 48.00 Matches: 10
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 90.91% Mismatches: 0
Query Match: 97.96% Indels: 0
DB: 6 Gaps: 0

US-09-965-594-26 (1-11) x AR145217 (1-75)

Qy 1 GlySerValValIleValGlyArgIleValLeu 11
|||||
Db 13 GGTCTGTGTTATTGTTGTAGAAATTATTTA 45

RESULT 14
AR145221
LOCUS AR145221
DEFINITION Sequence 55 from patent US 6211338.
ACCESSION AR145221
VERSION AR145221.1 GI:15107088
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 75)
AUTHORS Malcolim,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.
TITLE Single-chain recombinant complexes of hepatitis C virus NS3
protease and NS4A cofactor peptide
JOURNAL Patent: US 6211338-A 55 03-APR-2001;
FEATURES Location/Qualifiers
source
BASE COUNT 16 a 12 c 16 g 31 t
ORIGIN

Alignment Scores:
Pred. No.: 0.0408 Length: 75
Score: 48.00 Matches: 10
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 90.91% Mismatches: 0
Query Match: 97.96% Indels: 0
DB: 6 Gaps: 0

US-09-965-594-26 (1-11) x AR145221 (1-75)

Qy 1 GlySerValValIleValGlyArgIleValLeu 11
|||||
Db 13 GGTCTGTGTTATTGTTGTAGAAATTATTTA 45

RESULT 15
AR145197
LOCUS AR145197
DEFINITION Sequence 26 from patent US 6211338.
ACCESSION AR145197
VERSION AR145197.1 GI:15107064
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 78)
AUTHORS Malcolim,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.
TITLE Single-chain recombinant complexes of hepatitis C virus NS3
protease and NS4A cofactor peptide
JOURNAL Patent: US 6211338-A 26 03-APR-2001;

```

FEATURES      Location/Qualifiers
source        1..78
BASE COUNT    18 a /organism="unknown"
ORIGIN        11 c 18 g 31 t

Alignment Scores:
Pred. No.:    0.0426      Length:    78
Score:        48.00      Matches:    10
Percent Similarity: 100.00%      Conservative: 1
Best Local Similarity: 90.91%      Mismatches: 0
Query Match:  97.96%      Indels:    0
DB:           6          Gaps:      0

US-09-965-594-26 (1-11) x AR145197 (1-78)

Qy      1 GlySerValValIleValGlyArgIleValLeu 11
       |||||||
Db      13 GGTTCGTGTTATTGTTGTAAGATTATTTTA 45

Search completed: August 31, 2003, 00:46:50
Job time : 151.976 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 30, 2003, 19:13:57 : Search time 10.2149 Seconds
(without alignments)
2906.924 Million cell updates/sec

Title: US-09-965-594-26
Perfect score: 49
Sequence: 1 GSVVIVGRIVL 11

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-DB=N_Geneseq_19Jun03 -QWFAST=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09965594 -QCGN_1_1_1412 -runat_29082003_151918_28302 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_19Jun03:.*
1: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	49	100.0	68	21	AAA73336	HCV NS4A-NS3 fusio
2	49	100.0	71	21	AAA73337	HCV NS4A-NS3 fusio
3	49	100.0	588	21	AAA73328	Hepatitis C virus
4	49	100.0	588	21	AAA73329	Hepatitis C virus
5	49	100.0	618	13	AAQ32477	HCV NS2-NS4 clone
6	49	100.0	12734	24	ABA95615	Chimeric BVDV/HCV
7	48	98.0	75	20	AAH80310	HCV NS4A-NS3 compl
8	48	98.0	75	20	AAH80314	HCV NS4A-NS3 compl
9	48	98.0	78	20	AAH80306	HCV NS4A-NS3 compl
10	48	98.0	78	20	AAH80290	HCV NS4A-NS3 compl
11	48	98.0	96	20	AAH80338	HCV NS4A-NS3 compl
12	48	98.0	96	20	AAH80340	HCV NS4A-NS3 compl
13	48	98.0	161	24	AB235821	Hepatitis C virus
14	48	98.0	161	24	ABX10064	HCV NS4A DNA fragm
15	48	98.0	161	24	ABV78245	Hepatitis C virus
16	48	98.0	161	24	ABL91786	HCV polynucleotide
17	48	98.0	189	15	AAQ58472	HCV peptide c14-1.
18	48	98.0	267	15	AAQ58473	HCV peptide c14-1.
19	48	98.0	279	17	AAT26969	HCV II chimeric ep
20	48	98.0	283	18	AAT49363	Hepatitis C virus
21	48	98.0	321	17	AAT26968	HCV I chimeric epi
22	48	98.0	342	13	AAQ23456	DNA encoding non-A
23	48	98.0	372	13	AAQ24561	NANBH peptide B.
24	48	98.0	403	13	AAQ25743	Non-A, Non-B Hepat
25	48	98.0	403	13	AAQ25753	Non-A, Non-B Hepat
26	48	98.0	582	15	AAQ62690	HCV antigen. Synt
27	48	98.0	585	21	AA250045	DNA encoding Hepat
28	48	98.0	586	13	AAQ26990	HCV gene 10. Hepa
29	48	98.0	588	14	AAV05564	DNA associated wit
30	48	98.0	612	25	ABX15706	Anti-viral synthe
31	48	98.0	648	20	AAH80362	HCV NS4A-NS3 compl
32	48	98.0	648	20	AAH80363	HCV NS4A-NS3 compl
33	48	98.0	648	20	AAH80365	HCV NS4A-NS3 compl
34	48	98.0	650	20	AAH80346	HCV NS4A-NS3 compl
35	48	98.0	650	20	AAH80347	HCV NS4A-NS3 compl
36	48	98.0	651	20	AAH80342	HCV NS4A-NS3 compl
37	48	98.0	651	20	AAH80343	HCV NS4A-NS3 compl
38	48	98.0	651	20	AAH80344	HCV NS4A-NS3 compl
39	48	98.0	651	20	AAH80345	HCV NS4A-NS3 compl
40	48	98.0	651	20	AAH80348	HCV NS4A-NS3 compl
41	48	98.0	651	20	AAH80349	HCV NS4A-NS3 compl
42	48	98.0	651	20	AAH80350	HCV NS4A-NS3 compl
43	48	98.0	651	20	AAH80351	HCV NS4A-NS3 compl
44	48	98.0	654	21	AAZ50043	DNA encoding hepat
45	48	98.0	669	13	AAQ27012	HK10. Hepatitis C

ALIGNMENTS

RESULT 1
AAA73336
ID AAA73336 standard; DNA; 68 BP.
XX AAA73336;
AC
XX
XX 19-DEC-2000 (first entry)
DT
XX
XX HCV NS4A-NS3 fusion protease oligonucleotide #1.
XX
XX Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW liver failure; liver cancer; primer; ss.
XX
XX Synthetic.
XX
XX WO200040707-A1.
PN
XX

PD 13-JUL-2000.
 XX
 PF 06-JAN-2000; 2000WO-US00345.
 XX
 PR 08-JAN-1999; 99US-0115271.
 XX
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX
 PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
 XX
 DR WPI; 2000-465976/40.
 XX
 XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 XX
 XX Example 2; Page 22; 66pp; English.
 XX
 CC The present sequence is one of two oligonucleotides coding for NS4A and
 CC linker segments which were used in the creation of a fusion molecule of
 CC the coding sequences for Hepatitis C virus (HCV) NS3 and NS4A protease
 CC enzymes. These proteins are both essential for the replication of the
 CC virus, acting to cleave its replicative proteins from the polyprotein
 CC produced from the HCV genome. Inhibitors of the two proteins should be
 CC effective as antiviral treatments of HCV infection. This is useful as HCV
 CC can lead to chronic liver disease such as cirrhosis, liver failure and
 CC liver cancer. The present invention concerns a number of NS3 mutants and
 CC NS3-NS4A fusion proteins which can be used to identify inhibitors of this
 CC type, as well as enabling structural studies of the protease and
 CC protease:inhibitor complexes.
 XX
 SQ Sequence 68 BP; 20 A; 14 C; 17 G; 17 T; 0 other;

Alignment Scores:
 Pred. No.: 0.106 Length: 68
 Score: 49.00 Matches: 11
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 21 Gaps: 0

US-09-965-594-26 (1-11) x AAA73336 (1-68)
 QY 1 GlySerValIleValGlyArgIleValIleu 11
 DB 14 GGATCCGTTGTATCGTCGCCGTATAGTACTG 46

RESULT 2
 AAA73337/C
 ID AAA73337 standard; DNA; 71 BP.
 XX
 AC AAA73337;
 XX
 DT 19-DEC-2000 (first entry)
 XX
 DE HCV NS4A-NS3 fusion protease oligonucleotide #2.
 XX
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
 KW liver failure; liver cancer; primer; ss.
 OS Synthetic.
 XX
 PN WO200040707-A1.
 XX
 PD 13-JUL-2000.
 XX
 PF 06-JAN-2000; 2000WO-US00345.
 XX
 PR 08-JAN-1999; 99US-0115271.
 XX
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.

PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
 XX
 DR WPI; 2000-465976/40.
 XX
 PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
 PT amino acid, useful for screening inhibitors that may treat hepatitis C
 XX
 XX Example 2; Page 23; 66pp; English.
 XX
 CC The present sequence is one of two oligonucleotides coding for NS4A and
 CC linker segments which were used in the creation of a fusion molecule of
 CC the coding sequences for Hepatitis C virus (HCV) NS3 and NS4A protease
 CC enzymes. These proteins are both essential for the replication of the
 CC virus, acting to cleave its replicative proteins from the polyprotein
 CC produced from the HCV genome. Inhibitors of the two proteins should be
 CC effective as antiviral treatments of HCV infection. This is useful as HCV
 CC can lead to chronic liver disease such as cirrhosis, liver failure and
 CC liver cancer. The present invention concerns a number of NS3 mutants and
 CC NS3-NS4A fusion proteins which can be used to identify inhibitors of this
 CC type, as well as enabling structural studies of the protease and
 CC protease:inhibitor complexes.
 XX
 SQ Sequence 71 BP; 17 A; 18 C; 15 G; 21 T; 0 other;

Alignment Scores:
 Pred. No.: 0.111 Length: 71
 Score: 49.00 Matches: 11
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 21 Gaps: 0

US-09-965-594-26 (1-11) x AAA73337 (1-71)
 QY 1 GlySerValIleValGlyArgIleValIleu 11
 DB 60 GGATCCGTTGTATCGTCGCCGTATAGTACTG 28

RESULT 3
 AAA73328
 ID AAA73328 standard; DNA; 588 BP.
 XX
 AC AAA73328;
 XX
 DT 19-DEC-2000 (first entry)
 XX
 DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #1.
 XX
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
 KW liver failure; liver cancer; ds.
 XX
 OS Hepatitis C virus.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..588
 FT /*tag- a
 FT /product= "NS3-NS4A fusion protein"
 XX
 PN WO200040707-A1.
 XX
 PD 13-JUL-2000.
 XX
 PF 06-JAN-2000; 2000WO-US00345.
 XX
 PR 08-JAN-1999; 99US-0115271.
 XX
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX
 PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
 XX

DR WPI: 2000-465976/40.
DR P-PSDB; AAB15212.

XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
PT amino acid, useful for screening inhibitors that may treat hepatitis C
PT -

XX Claim 26; Fig 10; 66pp; English.

PS Disclosure; Fig 10; 66pp; English.

XX The present sequence is the coding sequence for a fusion protein created
CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
CC proteins are both essential for the replication of the virus, acting to
CC cleave its replicative proteins from the polyprotein produced from the
CC HCV genome. Inhibitors of the two proteins should be effective as
CC antiviral treatments of HCV infection. This is useful as HCV can lead to
CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
CC The present invention concerns a number of NS3 mutants and NS3-NS4A
CC fusion proteins which can be used to identify inhibitors of this type, as
CC well as enabling structural studies of the protease and
CC protease-inhibitor complexes.

XX Sequence 588 BP; 97 A; 183 C; 153 G; 155 T; 0 other;

Alignment Scores:
Pred. No.: 1.21 Length: 588
Score: 49.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 21 Gaps: 0

US-09-965-594-26 (1-11) x AAA73328 (1-588)

Qy 1 GlySerValIleValGlyArgIleValIleu 11
Db 13 GGTCGGTGTATCGTCGCCGTATAGTACTG 45

RESULT 4

AAA73329
ID AAA73329 standard; DNA; 588 BP.

XX AAA73329;

DT 19-DEC-2000 (first entry)

XX Hepatitis C virus NS4A-NS3 fusion protease coding sequence #2.

XX Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW liver failure; liver cancer; mutant; mutein; ds.

XX Hepatitis C virus.

OS Synthetic.

PH Key Location/Qualifiers

FT CDS 1..588

FT /*Lag- a /product= *NS4A-NS3 fusion protein #2*

XX WO2000040707-A1.

XX 13-JUL-2000.

XX 06-JAN-2000; 2000WO-US00345.

XX 08-JAN-1999; 99US-0115271.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;

XX WPI: 2000-465976/40.

DR P-PSDB; AAB15220.

XX

PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
PT amino acid, useful for screening inhibitors that may treat hepatitis C
PT -

XX Claim 26; Fig 12; 66pp; English.

XX The present sequence is the coding sequence for a mutated version of a
CC fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
CC protease enzymes. These proteins are both essential for the replication
CC of the virus, acting to cleave its replicative proteins from the
CC polyprotein produced from the HCV genome. Inhibitors of the two proteins
CC should be effective as antiviral treatments of HCV infection. This is
CC useful as HCV can lead to chronic liver disease such as cirrhosis, liver
CC failure and liver cancer. The present invention concerns a number of NS3
CC mutants and NS3-NS4A fusion proteins which can be used to identify
CC inhibitors of this type, as well as enabling structural studies of the
CC protease and protease-inhibitor complexes. The protein produced from this
CC sequence contains the alpha-helix0-1 variant.

XX Sequence 588 BP; 103 A; 180 C; 156 G; 149 T; 0 other;

Alignment Scores:
Pred. No.: 1.21 Length: 588
Score: 49.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 21 Gaps: 0

US-09-965-594-26 (1-11) x AAA73329 (1-588)

Qy 1 GlySerValIleValGlyArgIleValIleu 11
Db 13 GGATCGGTGTATCGTCGCCGTATAGTACTG 45

RESULT 5

AAA32477

ID AAA32477 standard; DNA; 618 BP.

XX AAA32477;

DT 25-MAR-2003 (updated)

DT 26-APR-1993 (first entry)

XX HCV NS2-NS4 clone N13-1.

XX Clone; polypeptide; NS2-NS4; Hepatitis C; Virus; HCV; serum; HC;
KW transcriptase; cDNA; primer; allele; ss.

XX Hepatitis C virus.

XX Key Location/Qualifiers

FT CDS 9..608

FT /*tag- a

XX EP518313-A2.

XX 16-DEC-1992.

XX 11-JUN-1992; 92EP-0109812.

XX 11-JUN-1991; 91JP-0139268.

XX 12-JUL-1991; 91JP-0172794.

XX 07-OCT-1991; 91JP-0287008.

XX 16-DEC-1991; 91JP-0332329.

XX 20-APR-1992; 92JP-0099957.

XX (MITU) MITSUBISHI KASEI CORP.

XX Hayashi N, Honda Y, Murakami T, Seki M, Takahashi K;

PI Teranishi Y;

XX WPI; 1992-417213/51.
 DR P-PSDB; AAR29846.
 XX
 PT New hepatitis C virus gene and its encoded protein - used for
 PT diagnosing and vaccinating against hepatitis C virus infections
 XX
 PS Disclosure; Page 125-26; 305pp; English.
 XX
 CC The sequences given in AAQ32472-82 and AAQ32442 are various clones which
 CC encode the NS2-NS4 regions of the hepatitis C virus (HCV) gene of
 CC the invention. These sequences were isolated from the serum of a
 CC patient suffering from hepatitis C (HC). The isolated RNA sequences
 CC were converted into cDNA using transcriptase in the presence of one
 CC of the primer sequences given in AAQ32553-64. The sequences were
 CC then amplified using primer pairs. The cDNA sequences isolated
 CC represent different alleles of the same region of the HCV gene.
 CC Sequence comparisons of these clones showed that it is possible for a
 CC patient to carry more than one HCV strain at one time. See also
 CC AAQ32436.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 618 BP; 135 A; 184 C; 176 G; 123 T; 0 other;
 Alignment Scores:
 Pred. No.: 1.28 Length: 618
 Score: 49.00 Matches: 11
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 13 Gaps: 0
 US-09-965-594-26 (1-11) x AAQ32477 (1-618)
 QY 1 GlySerValIleValGlyArgIleValIleu 11
 DB 351 GGCAGCGTGGTCATTTGGTGGCAGGATGCTTGTG 383
 RESULT 6
 AAQ95615
 ID ABA95615 standard; DNA; 12734 BP.
 AC ABA95615;
 XX
 XX 21-MAR-2002 (first entry)
 DE Chimeric BYDV/HCV NS3-wt sequence.
 DE
 DE
 KW Pestivirus; Npro; protease; NS3; screening; ds.
 XX
 OS Chimeric - Bovine viral diarrhea virus.
 OS Chimeric - Hepatitis C virus.
 XX
 PN US6326137-B1.
 XX
 PD 04-DEC-2001.
 XX
 PF 25-JUN-1999; 99US-0344456.
 XX
 PR 25-JUN-1999; 99US-0344456.
 XX
 PA (SCHE) SCHERING CORP.
 XX
 PI Hong Z, Lai VCH, Lau JYN;
 XX
 DR WPI; 2002-121103/16.
 XX
 CC Nucleic acid construct encoding chimeric Hepatitis C Virus (HCV)
 PT pestivirus genome where the Npro protease gene is replaced with NS3
 PT protease gene, useful for in vivo screening of compounds which inhibit
 PT HCV infection
 XX
 PS Example 2; Columns 17-28; 20pp; English.

XX The present invention relates to a nucleic acid construct encoding a
 CC chimeric Hepatitis C virus (HCV)-pestivirus genome. The construct
 CC comprises a pestivirus genome where a Npro pestivirus protease gene is
 CC replaced with a gene encoding a functional HCV NS3 protease. Furthermore,
 CC each junction site recognised by the Npro protease is replaced with a
 CC junction site recognised by the HCV NS3 protease. The construct is useful
 CC for screening compounds that inhibit HCV in vivo by inhibiting HCV
 CC protease, where screening may be in cell culture or in an animal model.
 CC The present sequence is a chimeric clone of BYDV (bovine viral diarrhea
 CC virus)/HCV NS3-wt, which was used to illustrate the present invention.
 XX
 SQ Sequence 12734 BP; 4032 A; 2604 C; 3295 G; 2803 T; 0 other;
 Alignment Scores:
 Pred. No.: 39.4 Length: 12734
 Score: 49.00 Matches: 11
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 24 Gaps: 0
 US-09-965-594-26 (1-11) x ABA95615 (1-12734)
 QY 1 GlySerValIleValGlyArgIleValIleu 11
 DB 413 GGTAGTGTGTATTGTGTGTAATGTTTGA 445
 RESULT 7
 AAQ80310
 ID AAX80310 standard; DNA; 75 BP.
 XX
 AC AAX80310;
 XX
 DT 07-SEP-1999 (first entry)
 XX
 DE HCV NS4A-NS3 complex construction primer SEQ ID NO:48.
 XX
 KW HCV; hepatitis C virus; single chain recombinant complex; linker;
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
 KW hydrophobic domain; covalent complex; detection; inhibitor; primer; ss.
 XX
 OS Synthetic.
 OS Hepatitis C virus.
 XX
 PN W0928482-A2.
 XX
 PD 10-JUN-1999.
 XX
 PF 24-NOV-1998; 98WO-US24528.
 XX
 PR 28-JUL-1998; 98US-0094331.
 PR 28-NOV-1997; 97US-0067315.
 XX
 PA (SCHE) SCHERING CORP.
 XX
 PI Malcolm BA, Taremi SS, Weber PC, Yao N;
 XX
 DR WPI; 1999-385385/32.
 XX
 PT New hepatitis C virus covalent complexes
 XX
 PS Example 1; Page 32; 21pp; English.
 XX
 CC The present invention describes a covalent hepatitis C virus (HCV)
 CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
 CC to the amino terminus of the HCV NS3 protease domain. The covalent
 CC NS4A-NS3 complexes are useful for structural determination and
 CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.
 CC They can also be used for detecting inhibitors of the protease activity,
 CC the helicase activity and the ATPase activity of NS3. The covalent

CC NS4A-NS3 complexes are more soluble, stable and active than the
CC non-covalent protease-peptide complexes previously available. The
CC present sequence represents a primer used in the construction of the
CC HCV NS4A-NS3 complexes.
XX
SQ Sequence 75 BP; 16 A; 14 C; 15 G; 30 T; 0 other;
Alignment Scores:
Pred. No.: 0.186 Length: 75
Score: 48.00 Matches: 10
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 90.91% Mismatches: 0
Query Match: 97.96% Indels: 0
DB: 20 Gaps: 0
US-09-965-594-26 (1-11) x AAX80310 (1-75)
OY 1 GlySerValIleValGlyArgIleValIleu 11
DB 13 GGTCTGTGTTATTGTTGGTAGAATTATTTA 45
RESULT 8
AAX80314
ID AAX80314 standard; DNA; 75 BP.
XX
AC AAX80314;
XX
DT 07-SEP-1999 (first entry)
XX
DE HCV NS4A-NS3 complex construction primer SEQ ID NO:55.
XX
KW HCV; hepatitis C virus; single chain recombinant complex; linker;
KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
KW hydrophobic domain; covalent complex; detection; inhibitor; primer; ss.
XX
OS Synthetic.
OS Hepatitis C virus.
XX
PN WO9928482-A2.
XX
PD 10-JUN-1999.
XX
PF 24-NOV-1998; 98WO-US24528.
XX
PR 28-JUL-1998; 98US-0094331.
PR 28-NOV-1997; 97US-0067315.
XX
PA (SCHE) SCHERING CORP.
XX
PI Malcolm BA, Taremi SS, Weber PC, Yao N;
XX
DR WPI; 1999-385385/32.
XX
PT New hepatitis C virus covalent complexes
XX
PS Example 1; Page 33; 21pp; English.
XX
CC The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The covalent
CC NS4A-NS3 complexes are useful for structural determination and
CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.
CC They can also be used for detecting inhibitors of the protease activity,
CC the helicase activity and the ATPase activity of NS3. The covalent
CC NS4A-NS3 complexes are more soluble, stable and active than the
CC non-covalent protease-peptide complexes previously available. The
CC present sequence represents a primer used in the construction of the
CC HCV NS4A-NS3 complexes.
XX
SQ Sequence 75 BP; 16 A; 12 C; 16 G; 31 T; 0 other;

Alignment Scores:
Pred. No.: 0.186 Length: 75
Score: 48.00 Matches: 10
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 90.91% Mismatches: 0
Query Match: 97.96% Indels: 0
DB: 20 Gaps: 0
US-09-965-594-26 (1-11) x AAX80314 (1-75)
OY 1 GlySerValIleValGlyArgIleValIleu 11
DB 13 GGTCTGTGTTATTGTTGGTAGAATTATTTA 45
RESULT 9
AAX80306
ID AAX80306 standard; DNA; 78 BP.
XX
AC AAX80306;
XX
DT 07-SEP-1999 (first entry)
XX
DE HCV NS4A-NS3 complex construction primer SEQ ID NO:42.
XX
KW HCV; hepatitis C virus; single chain recombinant complex; linker;
KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
KW hydrophobic domain; covalent complex; detection; inhibitor; primer; ss.
XX
OS Synthetic.
OS Hepatitis C virus.
XX
PN WO9928482-A2.
XX
PD 10-JUN-1999.
XX
PF 24-NOV-1998; 98WO-US24528.
XX
PR 28-JUL-1998; 98US-0094331.
PR 28-NOV-1997; 97US-0067315.
XX
PA (SCHE) SCHERING CORP.
XX
PI Malcolm BA, Taremi SS, Weber PC, Yao N;
XX
DR WPI; 1999-385385/32.
XX
PT New hepatitis C virus covalent complexes
XX
PS Example 1; Page 31; 21pp; English.
XX
CC The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The covalent
CC NS4A-NS3 complexes are useful for structural determination and
CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.
CC They can also be used for detecting inhibitors of the protease activity,
CC the helicase activity and the ATPase activity of NS3. The covalent
CC NS4A-NS3 complexes are more soluble, stable and active than the
CC non-covalent protease-peptide complexes previously available. The
CC present sequence represents a primer used in the construction of the
CC HCV NS4A-NS3 complexes.
XX
SQ Sequence 78 BP; 16 A; 14 C; 17 G; 31 T; 0 other;
Alignment Scores:
Pred. No.: 0.195 Length: 78
Score: 48.00 Matches: 10
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 90.91% Mismatches: 0
Query Match: 97.96% Indels: 0
DB: 20 Gaps: 0

US-09-965-594-26 (1-11) x AAX80306 (1-78)

QY 1 GlySerValIleValGlyArgIleValLeu 11
 DB 13 GGTCTGTTGTTATTGTTGTTAGAAATTATTTTA 45

RESULT 10

AAX80290
 ID AAX80290 standard; DNA; 78 BP.

XX AAX80290;

XX 07-SEP-1999 (first entry)

XX HCV NS4A-NS3 complex construction primer SEQ ID NO:26.

XX HCV; hepatitis C virus; single chain recombinant complex; linker;
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
 KW hydrophobic domain; covalent complex; detection; inhibitor; primer; ss.

XX Synthetic.

OS Hepatitis C virus.

XX WO9928482-A2.

XX 10-JUN-1999.

XX 24-NOV-1998; 98WO-US24528.

XX 28-JUL-1998; 98US-0094331.

XX 28-NOV-1997; 97US-0067315.

XX (SCHE) SCHERING CORP.

XX Malcolm BA, Taremi SS, Weber PC, Yao N;

XX WPI; 1999-385385/32.

XX New hepatitis C virus covalent complexes

XX Example 1; Page 26; 21lpp; English.

XX The present invention describes a covalent hepatitis C virus (HCV)
 CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
 CC to the amino terminus of the HCV NS3 protease domain. The covalent
 CC NS4A-NS3 complexes are useful for structural determination and
 CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.
 CC They can also be used for detecting inhibitors of the protease activity,
 CC the helicase activity and the ATPase activity of NS3. The covalent
 CC NS4A-NS3 complexes are more soluble, stable and active than the
 CC non-covalent protease-peptide complexes previously available. The
 CC present sequence represents a primer used in the construction of the
 CC HCV NS4A-NS3 complexes.

XX Sequence 78 BP; 18 A; 11 C; 18 G; 31 T; 0 other;

Alignment Scores:
 Pred. No.: 0.195 Length: 78
 Score: 48.00 Matches: 10
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 90.91% Mismatches: 0
 Query Match: 97.96% Indels: 0
 DB: 20 Gaps: 0

US-09-965-594-26 (1-11) x AAX80290 (1-78)

QY 1 GlySerValIleValGlyArgIleValLeu 11
 DB 13 GGTCTGTTGTTATTGTTGTTAGAAATTATTTTA 45

RESULT 11

AAX80338
 ID AAX80338 standard; DNA; 96 BP.

XX AAX80338;

XX 07-SEP-1999 (first entry)

XX HCV NS4A-NS3 complex construction primer SEQ ID NO:87.

XX HCV; hepatitis C virus; single chain recombinant complex; linker;
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
 KW hydrophobic domain; covalent complex; detection; inhibitor; primer; ss.

XX Synthetic.

OS Hepatitis C virus.

XX WO9928482-A2.

XX 10-JUN-1999.

XX 24-NOV-1998; 98WO-US24528.

XX 28-JUL-1998; 98US-0094331.

XX 28-NOV-1997; 97US-0067315.

XX (SCHE) SCHERING CORP.

XX Malcolm BA, Taremi SS, Weber PC, Yao N;

XX WPI; 1999-385385/32.

XX New hepatitis C virus covalent complexes

XX Example 2; Page 43; 21lpp; English.

XX The present invention describes a covalent hepatitis C virus (HCV)
 CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
 CC to the amino terminus of the HCV NS3 protease domain. The covalent
 CC NS4A-NS3 complexes are useful for structural determination and
 CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.
 CC They can also be used for detecting inhibitors of the protease activity,
 CC the helicase activity and the ATPase activity of NS3. The covalent
 CC NS4A-NS3 complexes are more soluble, stable and active than the
 CC non-covalent protease-peptide complexes previously available. The
 CC present sequence represents a primer used in the construction of the
 CC HCV NS4A-NS3 complexes.

XX Sequence 96 BP; 21 A; 17 C; 21 G; 37 T; 0 other;

Alignment Scores:
 Pred. No.: 0.246 Length: 96
 Score: 48.00 Matches: 10
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 90.91% Mismatches: 0
 Query Match: 97.96% Indels: 0
 DB: 20 Gaps: 0

US-09-965-594-26 (1-11) x AAX80338 (1-96)

QY 1 GlySerValIleValGlyArgIleValLeu 11
 DB 31 GGTCTGTTGTTATTGTTGTTAGAAATTATTTTA 63

RESULT 12

AAX80340
 ID AAX80340 standard; DNA; 96 BP.

XX AAX80340;

XX 07-SEP-1999 (first entry)

XX HCV NS4A-NS3 complex construction primer SEQ ID NO:89.
DE HCV; hepatitis C virus; single chain recombinant complex; linker;
XX NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
KW hydrophobic domain; covalent complex; detection; inhibitor; primer; ss.
KW Synthetic.
OS Hepatitis C virus.
OS WO9928482-A2.
XX 10-JUN-1999.
XX 24-NOV-1998; 98WO-US24528.
XX 28-JUL-1998; 98US-0094331.
PR 28-NOV-1997; 97US-0067315.
XX (SCHE) SCHERING CORP.
XX Malcolm BA, Taremi SS, Weber PC, Yao N;
PI WPI; 1999-385385/32.
XX New hepatitis C virus covalent complexes
XX Example 2; Page 44; 21pp; English.
XX The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The covalent
CC NS4A-NS3 complexes are useful for structural determination and
CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.
CC They can also be used for detecting inhibitors of the protease activity,
CC the helicase activity and the ATPase activity of NS3. The covalent
CC NS4A-NS3 complexes are more soluble, stable and active than the
CC non-covalent protease-peptide complexes previously available. The
CC present sequence represents a primer used in the construction of the
CC HCV NS4A-NS3 complexes.
XX SQ Sequence 96 BP; 21 A; 17 C; 21 G; 37 T; 0 other;

Alignment Scores:
Pred. No.: 0.246 Length: 96
Score: 48.00 Matches: 10
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 90.91% Mismatches: 0
Query Match: 97.96% Indels: 0
DB: 20 Gaps: 0

US-09-965-594-26 (1-11) x AAX80340 (1-96)

QY 1 GlySerValIleValGlyArgIleValLeu 11
DB 31 GGTCGCTGTTATTGCTGCTAGAAATATTTTA 63

RESULT 13
ABZ35821
ID ABZ35821 standard; DNA; 161 BP.
XX ABZ35821;
XX 07-FEB-2003 (first entry)
XX Hepatitis C virus NS4A polynucleotide SEQ ID NO 129.
XX Double stranded RNA; dsRNA; RNAi; RNA inhibition; cytostatic; virucide;
KW protozoicide; gene expression; antisense; tumour; infection; Plasmodium;
KW virus; viroid; anti-GFP; human; HTV; human immunodeficiency virus;
KW Hepatitis C virus; human papilloma virus; gene; ds.

XX Hepatitis C Virus.
OS DE10100588-A1.
XX 18-JUL-2002.
XX 09-JAN-2001; 2001DE-1000588.
XX 09-JAN-2001; 2001DE-1000588.
XX (RIBO-) RIBOPHARMA AG.
XX Kreutzer R, Limmer S, Rost S, Hadwiger P;
PI WPI; 2002-683450/74.
XX Inhibiting expression of target genes, useful e.g. for treating tumors,
PT by introducing into cells two double-stranded RNAs that are
PT complementary to the target
XX Claim 13; Page 87; 100pp; German.
XX The invention relates to inhibiting expression of a target gene in a cell
CC by introducing at least two oligoribonucleotides (dsRNAI and II), both
CC with a double-stranded (ds) structure of at most 49 sequential nucleotide
CC pairs. At least part of one strand (S1, S2) of the ds structures in each
CC of dsRNAI and II are complementary to regions in the target gene. The
CC method uses antisense inhibition of gene expression using double stranded
CC RNA inhibition (RNAi). The method is particularly used to treat tumours
CC or infections, especially by Plasmodium or viruses/viroids (pathogenic on
CC humans, animals or plants). The method provides more effective inhibition
CC of expression than known methods using a single dsRNA, even at very low
CC concentrations. When dsRNA has at least one unpaired nucleotide at the
CC end, stability (and thus effective concentration in the cell) is
CC improved and efficiency can be increased further by pretreating the cells
CC with interferon. The present sequence is that of a target DNA of the
CC invention.
XX SQ Sequence 161 BP; 32 A; 40 C; 55 G; 34 T; 0 other;

Alignment Scores:
Pred. No.: 0.442 Length: 161
Score: 48.00 Matches: 10
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 90.91% Mismatches: 0
Query Match: 97.96% Indels: 0
DB: 24 Gaps: 0

US-09-965-594-26 (1-11) x ABZ35821 (1-161)

QY 1 GlySerValIleValGlyArgIleValLeu 11
DB 60 GGCAGCGTGTCTATTGTGGCAGGATCATCTTG 92

RESULT 14
ABX10064
ID ABX10064 standard; DNA; 161 BP.
XX ABX10064;
XX 23-JAN-2003 (first entry)
XX HCV NS4A DNA fragment SEQ ID 129.
XX Oligoribonucleotide; interferon; oncogene; cytokine; Id; developmental;
KW prion; inhibition; ds.
XX Hepatitis C virus.
XX DE10100587-C1.
XX 21-NOV-2002.

XX 09-JAN-2001; 2001DE-1000587.
 XX 09-JAN-2001; 2001DE-1000587.
 XX (RIBO-) RIBOPHARMA AG.

XX Kreutzer R, Limmer S, Rost S, Hadwiger P;
 XX WPI; 2002-742209/81.

XX Inhibiting expression of target genes, e.g. oncogenes, in cells, by
 PT introduction of complementary double-stranded oligoribonucleotide,
 PT after treating the cell with interferon

XX Disclosure; Page 92; 98pp; German.

XX This invention describes a novel method for inhibiting expression of a
 CC target gene by introducing into the cell that contains the target gene
 CC at least one oligoribonucleotide (dsRNA) that has a double-stranded
 CC (ds) structure of not more than 49 consecutive nucleotides (nt), where
 CC at least a segment of one strand of the ds structure is complementary
 CC with the target gene and the cells are treated with interferon before
 CC introduction of dsRNA. The method is used to inhibit expression of
 CC target genes, particularly oncogenes, cytokine genes, Id (not defined)
 CC protein genes, developmental or prion genes, or genes expressed in
 CC pathogenic organisms (particularly plasmidia) or in viruses or viroids
 CC (pathogenic in humans, animals or plants). Treating the cells with
 CC interferon greatly increases the extent to which dsRNA can inhibit
 CC expression of the target genes, and the effect is even greater when dsRNA
 CC are modified to increase their stability. ABX09936-ABX10075 represent
 CC gene fragments used to illustrate the method of the invention.

XX Sequence 161 BP; 32 A; 40 C; 55 G; 34 T; 0 other;

Alignment Scores: 0.442 Length: 161
 Pred. No.: 48.00 Matches: 10
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 90.91% Mismatches: 0
 Query Match: 97.96% Indels: 0
 DB: 24 Gaps: 0

US-09-965-594-26 (1-11) x ABX10064 (1-161)

Qy 1 GlySerValValIleValGlyArgIleValIleu 11
 |||||
 Db 60 GGCACGCGTGTCAATTGTGGCAGCATCTTG 92

RESULT 15

ABV78245
 ID ABV78245 standard; DNA; 161 BP.

XX AC ABV78245;

XX DT 15-NOV-2002 (first entry)

XX Hepatitis C virus NS4A DNA SEQ ID NO 129.

DE RNA inhibition; dsRNA; gene expression inhibitor; oncogene; cytostatic;
 KW virucide; protozoacide; gene; ds.

OS Hepatitis C virus.

XX WO200255693-A2.

XX 18-JUL-2002.

XX PF 09-JAN-2002; 2002WO-EP00152.

XX 09-JAN-2001; 2001DE-1000586.

PR 26-OCT-2001; 2001DE-1055280.

PR 29-NOV-2001; 2001DE-1056411.

PR 07-DEC-2001; 2001DE-1060151.

XX (RIBO-) RIBOPHARMA AG.

XX Kreutzer R, Limmer S, Rost S, Hadwiger P;

XX WPI; 2002-590671/63.

XX Inhibiting expression of target gene, useful e.g. for inhibiting
 PT oncogenes, by administering double-stranded RNA complementary to the
 PT target and having an overhang

XX Claim 10; Page 190; 203pp; German.

XX The invention relates to inhibiting expression of a target gene (I) in a
 CC cell by introducing an inhibitory RNA (dsRNA) having a double-stranded
 CC structure of at most 49 consecutive bases. At least part of one strand
 CC (asI) of dsRNA is complementary to (I) and at least one end of dsRNA
 CC has an overhang of 1-4 nucleotides. The method is used to inhibit the
 CC expression of a wide range of genes, e.g. oncogenes, cytokine genes etc.
 CC in humans, also genes in plasmidium or in viruses or viroids that are
 CC pathogenic for humans, animals or plants. Introducing an overhang into
 CC dsRNA greatly increases effectiveness for inhibiting gene expression,
 CC both in vivo and in vitro and also increases stability and thus the
 CC effective concentration inside the cell. The present sequence is that of
 CC a gene related to the invention.

XX Sequence 161 BP; 32 A; 40 C; 55 G; 34 T; 0 other;

Alignment Scores: 0.442 Length: 161
 Pred. No.: 48.00 Matches: 10
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 90.91% Mismatches: 0
 Query Match: 97.96% Indels: 0
 DB: 24 Gaps: 0

US-09-965-594-26 (1-11) x ABV78245 (1-161)

Qy 1 GlySerValValIleValGlyArgIleValIleu 11
 |||||
 Db 60 GGCACGCGTGTCAATTGTGGCAGCATCTTG 92

Search completed: August 30, 2003, 19:48:23

Job time : 12.2149 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: August 30, 2003, 19:20:43 ; Search time 106.667 Seconds
(without alignments)
2506.388 Million cell updates/sec

Title: US-09-965-594-26
Perfect score: 49
Sequence: 1 GSWVIVGRIVL 11

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-Q/cgn2_1/USPTO.spool/US09965594/runat_29082003_151919_28322/app.query.fasta_1.2872
-DB-EST -OPTM-fastap -SUFFIX-rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS-bits -START=1 -END=1 -MATRIX-blosum62 -TRANS-human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE-pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM-ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09965594.ecgn_1_1_12630/runat_29082003_151919_28322 -NCPU=3
-NO_MMALP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estml:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pin:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
1	45	91.8	234	9	AW580055
2	43	87.8	618	14	CAT28563
3	43	87.8	664	13	BM300987
4	43	87.8	709	13	BM306038
5	43	87.8	725	13	BM176544
6	42	85.7	150	10	BE069681
7	42	85.7	646	10	BE069684
8	42	85.7	658	13	BQ121075
9	42	85.7	777	10	BE060305
10	42	85.7	1035	13	BU718639
11	41	83.7	222	9	AU183349
12	41	83.7	487	12	BI882963
13	41	83.7	502	10	BE556909
14	41	83.7	564	14	CB417286
15	41	83.7	584	12	BI981053
16	41	83.7	597	12	BI840872
17	41	83.7	603	12	BM185756
18	41	83.7	607	13	BQ093551
19	41	83.7	620	12	BI982002
20	41	83.7	629	10	BE016243
21	41	83.7	641	10	BE016238
22	41	83.7	641	10	BE201405
23	41	83.7	660	14	CD392599
24	41	83.7	676	9	AW019436
25	41	83.7	796	9	AF122168
26	41	83.7	882	13	BQ734390
27	41	83.7	886	12	BI759235
28	41	83.7	1569	29	CC250893
29	40	81.6	315	9	AV211292
30	40	81.6	398	29	CNS03X0L
31	40	81.6	540	10	BG403954
32	40	81.6	667	14	BZ004158
33	40	81.6	693	28	BZ004158
34	40	81.6	771	13	BQ785063
35	40	81.6	791	29	AG096061
36	40	81.6	1280	29	AG092509
37	40	81.6	1657	11	BG708760
38	40	81.6	1976	11	BC035952
39	39	79.6	116	10	BE069569
40	39	79.6	135	10	BE069567
41	39	79.6	170	10	BE071133
42	39	79.6	178	10	BE069730
43	39	79.6	195	10	BE071108
44	39	79.6	234	9	AV012254
45	39	79.6	251	10	BB441428

ALIGNMENTS

RESULT 1
AW580055
LOCUS RC1-HT0375-130100-011-905 HT0375 Homo sapiens cDNA, mRNA linear EST 16-MAR-2000
DEFINITION AW580055
ACCESSION AW580055
VERSION AW580055.1 GI:7255104
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 234)
AUTHORS HCCP <http://www.ludwig.org.br/ORESTES>.

TITLE The FAPESP/LICR Human Cancer Genome Project
JOURNAL Unpublished
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l1-RC1&t2-RC1-HT0375-130100-011-g05&t3=2000-01-13&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 28
High quality sequence stop: 216.

FEATURES

source
1..234
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="HT0375"

/note="organ: head_neck; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 52 a 35 c 76 g 71 t

ORIGIN

Alignment Scores: 26.4 Length: 234
Pred. No.: 45.00 Matches: 8
Score: 100.00% Conservative: 3
Percent Similarity: 72.73% Mismatches: 0
Best Local Similarity: 91.84% Indels: 0
Query Match: 9 Gaps: 0
DB:

US-09-965-594-26 (1-11) x AW580055 (1-234)

QY 1 GlySerValIleValGlyArgIleValLeu 11
|||||:|||||:|||||:|||||:|||||

Db 75 GGTAGTATTATTAATTCGGGACGATTAGTTTA 107

RESULT 2

CA728563/c

LOCUS 618 bp mRNA linear EST 26-NOV-2002
DEFINITION wdilc.pk004.f22 wdilc Triticum aestivum cDNA clone wdilc.pk004.f22
5' end mRNA sequence.

ACCESSION CA728563.1 GI:25450552

VERSION EST.

KEYWORDS Triticum aestivum (bread wheat)

SOURCE

ORGANISM

Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Triticeae; Triticum.

1 (bases 1 to 618)

Tingey,S.V., Powell,W., Dolan,M., Hainey,C., Yuan,Z.,

Miao,G., Caraher,N. and Hanafey,M.K.

DuPont Wheat cDNA Sequence

Unpublished

Contact: Scott V. Tingey

Crop Genetics

E. I. DuPont de Nemours and Company

1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA

Tel: 302-631-2602

Fax: 302-631-2607

Email: Scott.V.Tingey@USA.dupont.com

FEATURES

Seq primer: M13.
Location/Qualifiers
1..618
/organism="Triticum aestivum"
/mol_type="mRNA"
/db_xref="taxon:4565"
/clone_lib="wdilc.pk004.f22"
/tissue_type="inflorescence"
/lab_host="DH10B"
/clone_lib="wdilc"
/note="Vector: pBluescript SK+; Site_1: EcoRI; Site_2:
XhoI; Wheat (Triticum aestivum, Hi Line) developing
inflorescence +/- 4 cm"

BASE COUNT 140 a 194 c 168 g 110 t 6 others

ORIGIN

Alignment Scores: 235 Length: 618
Pred. No.: 43.00 Matches: 8
Score: 90.91% Conservative: 2
Percent Similarity: 72.73% Mismatches: 1
Best Local Similarity: 87.76% Indels: 0
Query Match: 14 Gaps: 0
DB:

US-09-965-594-26 (1-11) x CA728563 (1-618)

QY 1 GlySerValIleValGlyArgIleValLeu 11

|||||:|||||:|||||:|||||:|||||

Db 113 GCAGGGGTGCTGCTGGGACGAGTGCTCTT 81

RESULT 3

BW300987

LOCUS

DEFINITION BW300987 664 bp mRNA linear EST 11-NOV-2002
Intestinalis cDNA clone cinc020g06 5', mRNA sequence.

ACCESSION BW300987.1 GI:24881598

VERSION EST.

KEYWORDS

SOURCE

ORGANISM

Ciona intestinalis
Ciona intestinalis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
Phlebobranchia; Clonidae; Ciona.

1 (bases 1 to 664)

Satou,Y., Shin-i,T., Kohara,Y. and Satoh,N.

Expressed genes in Ciona intestinalis (2002c)

Unpublished

Contact: Nori Satoh

Department of Zoology

Kyoto University

Sakyo-ku, Kyoto, Kyoto 606-8502, Japan

Tel: 81-75-753-4081

Fax: 81-75-705-1113

Email: satohe@scidian.zool.kyoto-u.ac.jp.

FEATURES

source

1..664
Location/Qualifiers
/organism="Ciona intestinalis"
/mol_type="mRNA"
/db_xref="taxon:7719"
/clone_lib="cinc020g06"
/tissue_type="neural complex"
/clone_lib="Nori Satoh unpublished cDNA library, neural
complex"

BASE COUNT 201 a 145 c 135 g 183 t

ORIGIN

Alignment Scores: 259 Length: 664
Pred. No.: 43.00 Matches: 9
Score: 100.00% Conservative: 1
Percent Similarity: 90.00% Mismatches: 0
Best Local Similarity: 87.76% Indels: 0
Query Match: 13 Gaps: 0
DB:

US-09-965-594-26 (1-11) x BW300987 (1-664)

Qy 1 GlySerValIleValGlyArgIleVal 10
 |||||
 Db 418 GGAAGTGTGTGTGTGTGTGGAAGATTGTT 447

RESULT 4

BW306038

LOCUS

DEFINITION BW306038 Nori Satoh unpublished cDNA library, heart EST 11-NOV-2002
 intestinalis cDNA clone ciht012f15 5', mRNA sequence.

ACCESSION BW306038

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Expressed genes in Ciona intestinalis (2002c)

Unpublished

Contact: Nori Satoh

Department of Zoology

Kyoto University

Sakyo-ku, Kyoto, Kyoto 606-8502, Japan

Tel: 81-75-753-4081

Fax: 81-75-705-1113

Email: satoheascidian.zool.kyoto-u.ac.jp.

Location/Qualifiers

1..709

/organism="Ciona intestinalis"

/mol_type="mRNA"

/db_xref="taxon:7719"

/clone="ciht012f15"

/tissue_type="heart"

/clone_lib="Nori Satoh unpublished cDNA library, heart"

BASE COUNT 199 a 156 c 155 g 199 t

ORIGIN

Alignment Scores:

Pred. No.: 283

Score: 43.00

Percent Similarity: 100.00%

Best Local Similarity: 90.00%

Query Match: 87.76%

DB: 13

Length: 709

Matches: 9

Conservative: 1

Mismatches: 0

Indels: 0

Gaps: 0

US-09-965-594-26 (1-11) x BW306038 (1-709)

Qy 1 GlySerValIleValGlyArgIleVal 10

|||||

Db 307 GGAAGTGTGTGTGTGTGGAAGATTGTT 336

RESULT 5

BW176544/c

LOCUS

DEFINITION BW176544 Nori Satoh unpublished cDNA library, heart EST 04-NOV-2002
 intestinalis cDNA clone rciht012f15 3', mRNA sequence.

ACCESSION BW176544

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Expressed genes in Ciona intestinalis (2002c)

Unpublished

Contact: Nori Satoh

Department of Zoology

Kyoto University

Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
 Tel: 81-75-753-4081
 Fax: 81-75-705-1113
 Email: satoheascidian.zool.kyoto-u.ac.jp.

FEATURES
 source

1..725

/organism="Ciona intestinalis"

/mol_type="mRNA"

/db_xref="taxon:7719"

/clone="rciht012f15"

/tissue_type="heart"

/clone_lib="Nori Satoh unpublished cDNA library, heart"

BASE COUNT 195 a 168 c 155 g 207 t

ORIGIN

Alignment Scores:

Pred. No.: 292

Score: 43.00

Percent Similarity: 100.00%

Best Local Similarity: 90.00%

Query Match: 87.76%

DB: 13

Length: 725

Matches: 9

Conservative: 1

Mismatches: 0

Indels: 0

Gaps: 0

US-09-965-594-26 (1-11) x BW176544 (1-725)

Qy 1 GlySerValIleValGlyArgIleVal 10

|||||

Db 546 GGAAGTGTGTGTGTGGAAGATTGTT 517

RESULT 6

BE069681

LOCUS

DEFINITION RC2-BT0389-120400-014-c02 BT0389 Homo sapiens cDNA, mRNA sequence.

ACCESSION BE069681

VERSION BE069681.1

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 150)

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,

Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,

Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,

Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare

,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and

Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

10737800

COMMENT

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?l1=4t2-RC2-BT0389-120

400-014-c02&t3=2000-04-12&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 3

High quality sequence stop: 150.

Location/Qualifiers

1..150

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/dev_stage="Adult"

```

/clone.lib="BT0389"
/Note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT      38 a      18 c      45 g      49 t
ORIGIN
Alignment Scores:
Pred. No.:      55.3      Length:      150
Score:          42.00     Matches:      8
Percent Similarity: 90.91% Conservative: 2
Best Local Similarity: 72.73% Mismatches: 1
Query Match:     85.71% Indels: 0
DB:             10      Gaps: 0

US-09-965-594-26 (1-11) x BE069681 (1-150)
QY      1 GlySerValIleValGlyArgIleValLeu 11
Db      12 GGTAGTACTAATAATTGTGGGAGATTAGTTT 44

RESULT 7
LOCUS      BG590884               646 bp      mRNA      linear      EST 07-MAR-2003
DEFINITION EST498726 P. infestans-challenged leaf Solanum tuberosum cDNA clone
ACCESSION BPL16H5 5' sequence, mRNA sequence.
VERSION    BG590884.1 GI:13609024
KEYWORDS   EST.
SOURCE     Solanum tuberosum (potato)
ORGANISM   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 646)
AUTHORS    Zhang,P., Hernandez,M., Tornqvist,C.-E, Wirtz,U., Loukianov,A.,
Rangel,P., Haberland,G.T., Cho,J., Chiemingo,A., Bougri,O., Buell
,C.R., Ronning,C.M., Helgeson,J. and Baker,B.
TITLE      Generation of ESTs from Potato Leaves Challenged with Phytophthora
JOURNAL    Infestans, Incompatible Reaction
COMMENT    Unpublished
Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato-array@tigr.org
This clone can be obtained from the University of Arizona Genomics
Institute. Orders can be made through URL:
http://genome.arizona.edu/orders/
Seq primer: M13F-R.
FEATURES             Location/Qualifiers
     source           1..646
                     /organism="Solanum tuberosum"
                     /mol_type="mRNA"
                     /cultivar="Kennebec"
                     /db_xref="taxon:4113"
                     /clone="BPL16H5"
                     /tissue_type="leaf"
                     /dev_stage="6 week old"
                     /lab_host="SOLR"
                     /clone_lib="P. infestans-challenged leaf"
     note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Whole plants were challenged with 450,000
sporangia/ml P. infestans US-1(US 940501) in Biotron
(Madison, Wisconsin). Leaf tissue was collected at 1, 2,
5, 12, and 24 hours post-challenge and frozen in liquid
nitrogen immediately upon removal. Kennebec plants showed
no signs of HR. Katahdin plants (susceptible to P.
infestans US-1) were used as controls and showed
infection. NOTE: We cannot exclude the possibility that

```

```

this sequence is actually derived from Phytophthora rather
than potato."
BASE COUNT      159 a      110 c      163 g      214 t
ORIGIN
Alignment Scores:
Pred. No.:      390      Length:      646
Score:          42.00     Matches:      9
Percent Similarity: 90.91% Conservative: 1
Best Local Similarity: 81.82% Mismatches: 1
Query Match:     85.71% Indels: 0
DB:             10      Gaps: 0

US-09-965-594-26 (1-11) x BG590884 (1-646)
QY      1 GlySerValIleValGlyArgIleValLeu 11
Db      117 GGTCGTAGTATGATGGGAGATCATCATG 149

RESULT 8
LOCUS      BQ121075               658 bp      mRNA      linear      EST 07-MAR-2003
DEFINITION EST606651 mixed potato tissues Solanum tuberosum cDNA clone STMEV29
ACCESSION BQ121075               658 bp      mRNA      linear      EST 07-MAR-2003
VERSION    BQ121075.2 GI:21920306
KEYWORDS   EST.
SOURCE     Solanum tuberosum (potato)
ORGANISM   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 658)
AUTHORS    Buell,C.R., Hart,A., Baker,B., Tanksley,S., Fry,W., Smart,C.,
Restrepo,S., Griffiths,H., van der Hoeven,R., Tsai,J. and
Karamycheva,S.A.
TITLE      Generation of a set of potato cDNA clones for microarray analyses
JOURNAL    Unpublished
COMMENT    On Apr 17, 2002 this sequence version replaced gi:20173037.
Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato-array@tigr.org
This clone can be obtained from the University of Arizona Genomics
Institute. Orders can be made through URL:
http://genome.arizona.edu/orders/
Seq primer: T3.
FEATURES             Location/Qualifiers
     source           1..658
                     /organism="Solanum tuberosum"
                     /mol_type="mRNA"
                     /cultivar="Kennebec or Binjite"
                     /db_xref="taxon:4113"
                     /clone="STMEV29"
                     /tissue_type="mixed tissues"
                     /lab_host="SOLR"
                     /clone_lib="mixed potato tissues"
     note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Combination of untreated and Phytophthora
infestans-treated libraries of stolons, leaves, leaflets,
axillary buds of stem explants, petioles, germinating eyes
, tubers, or roots."
BASE COUNT      159 a      97 c      182 g      220 t
ORIGIN
Alignment Scores:
Pred. No.:      399      Length:      658
Score:          42.00     Matches:      9
Percent Similarity: 90.91% Conservative: 1
Best Local Similarity: 81.82% Mismatches: 1
Query Match:     85.71% Indels: 0
DB:             13      Gaps: 0

```


FEATURES

source Location/Qualifiers
 1. .222
 /organism="Cyprinus carpio"
 /mol_type="mRNA"
 /db_xref="taxon:7962"
 /clone="H7"
 /tissue_type="head kidney"
 /clone_lib="Cyprinus carpio head kidney stimulated by
 lipo-polysaccharide and concanavalin-A"
 /note="common name: common carp ; stimulated by
 lipo-polysaccharide and concanavalin-A"
 BASE COUNT 73 a 34 c 45 g 70 t
 ORIGIN

Alignment Scores:
 Pred. No.: 146 Length: 222
 Score: 41.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 4
 Best Local Similarity: 63.64% Mismatches: 0
 Query Match: 83.67% Indels: 0
 DB: 9 Gaps: 0

US-09-965-594-26 (1-11) x AU183349 (1-222)

QY 1 GlySerValValIleValGlyArgIleValLeu 11
 Db 42 GGGCAGTGGTTGTCATGGGAAGAGTTGTACTG 74

RESULT 12

BI882963/C
 LOCUS BI882963 487 bp mRNA linear EST 16-SEP-2002
 DEFINITION f01e01.xl zebrafish Research Genetics C32 fin Danio rerio cDNA
 clone IMAGE:4469065 3', mRNA sequence.
 ACCESSION BI882963
 VERSION BI882963.1 GI:16090234
 KEYWORDS EST.
 SOURCE Danio rerio (zebrafish)
 ORGANISM Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
 ; Cyprinidae; Danio.
 REFERENCE 1 (bases 1 to 487)
 AUTHORS Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy
 , S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood
 , K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B.,
 Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E.,
 Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R.
 and Wilson, R.

WashU Zebrafish EST Project 1998

TITLE Unpublished
 JOURNAL
 COMMENT Contact: Stephen L. Johnson
 Washington University School of Medicine.
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810

Email: zbrafish@wustl.edu
 CDNA Library Preparation: Ning Wu. CDNA Library Arrayed by: Steve
 Johnson. DNA Sequencing by: Washington University Genome Sequencing
 Center NOTE: This clone is available royalty-free through LNL;
 contact the IMAGE Consortium (info.llnl.gov) for further
 information.

Seq primer: T7 from Gibco

High quality sequence stop: 444.

Location/Qualifiers

FEATURES

source 1. .487
 /organism="Danio rerio"
 /mol_type="mRNA"
 /db_xref="taxon:7955"
 /clone="IMAGE:4469065"
 /tissue_type="Fin"
 /lab_host="GeneHogs (HS996, a phage-resistant isolate of
 DH10B)"
 /clone_lib="zebrafish Research Genetics C32 fin"

/note="vector: pT73D-Pac with a modified polylinker;
 Site 1: EcoRI; Site 2: NotI; 1st strand cDNA was prepared
 from zebrafish(C32) fin, and was then primed with a Not I
 - oligo(dT) primer. Double-stranded cDNA was ligated to
 Eco RI adaptors (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of the modified
 pT73 vector. Library is non-normalized. Library was
 constructed by Ning Wu. NOTE: This clone is available
 royalty-free through LNL; contact the IMAGE Consortium
 (info.llnl.gov) for further information"

BASE COUNT 144 a 115 c 91 g 137 t
 ORIGIN

Alignment Scores:
 Pred. No.: 417 Length: 487
 Score: 41.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 4
 Best Local Similarity: 63.64% Mismatches: 0
 Query Match: 83.67% Indels: 0
 DB: 12 Gaps: 0

US-09-965-594-26 (1-11) x BI882963 (1-487)

QY 1 GlySerValValIleValGlyArgIleValLeu 11
 Db 165 GGGCAGTGGTTGTCATGGGAAGAGTTGTACTG 133

RESULT 13

BE556909
 LOCUS BE556909 502 bp mRNA linear EST 30-AUG-2000
 DEFINITION fk92d02.y1 zebrafish Research Genetics C32 fin Danio rerio cDNA 5',
 mRNA sequence.
 ACCESSION BE556909
 VERSION BE556909.1 GI:9821399
 KEYWORDS EST.
 SOURCE Danio rerio (zebrafish)
 ORGANISM Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
 ; Cyprinidae; Danio.

REFERENCE 1 (bases 1 to 502)
 AUTHORS Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy
 , S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood
 , K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B.,
 Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E.,
 Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R.
 and Wilson, R.

WashU Zebrafish EST Project 1998

TITLE Unpublished

JOURNAL

COMMENT Contact: Stephen L. Johnson

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: zbrafish@wustl.edu

CDNA Library Preparation: Ning Wu. CDNA Library Arrayed by:
 Research Genetics. DNA Sequencing by: Washington University Genome
 Sequencing Center Clone Distribution: Research Genetics web
 address: <http://www.researchgenetics.com/>

Seq primer: T3 ET from Amersham

High quality sequence stop: 466.

Location/Qualifiers

1. .502

FEATURES

source /organism="Danio rerio"
 /mol_type="mRNA"
 /db_xref="taxon:7955"
 /tissue_type="Fin"
 /lab_host="GeneHogs (HS996, a phage-resistant isolate of
 DH10B)"
 /clone_lib="zebrafish Research Genetics C32 fin"
 /note="vector: pT73D-Pac with a modified polylinker;
 Site 1: EcoRI; Site 2: NotI; 1st strand cDNA was prepared
 from zebrafish(C32) fin, and was then primed with a Not I

- oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is non-normalized. Library was constructed by Ning Wu. NOTE: This clone is available royalty-free through LNL; contact the IMAGE Consortium (info.lnl.gov) for further information."

BASE COUNT 136 a 100 c 124 g 142 t

ORIGIN

Alignment Scores:

Pred. No.: 434 Length: 502
Score: 41.00 Matches: 7
Percent Similarity: 100.00% Conservative: 4
Best Local Similarity: 63.64% Mismatches: 0
Query Match: 83.67% Indels: 0
DB: 10 Gaps: 0

US-09-965-594-26 (1-11) x BE556909 (1-502)

QY 1 GlySerValIleValGlyArgIleValLeu 11

Db 355 GGGCAGTGGTGTCTCATGGGAAGTGTACTG 387

RESULT 14

LOCUS

DEFINITION CB417286 564 bp mRNA linear EST 27-MAR-2003
STR00763 gastrula stage cDNA library Danio rerio cDNA clone CB380
5' similar to CCAAT/enhancer binding protein beta, mRNA sequence.

ACCESSION

CB417286

VERSION

CB417286.1

KEYWORDS

EST.

SOURCE

ORGANISM

Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes

; Cyprinidae; Danio.

REFERENCE

1 (bases 1 to 564)

AUTHORS

Loppin,B., Steffan,T., Kempf,J., Heyer,V., Thisse,C. and Thisse,B.

Expression of the zebrafish genome during embryogenesis

Unpublished

JOURNAL

COMMENT

Contact: Thisse B

Institut de Genetique et de Biologie Moleculaire et Cellulaire

CNRS, INSERM, ULP

1, rue Laurent Fries, BP163, CU de Strasbourg, 67404 Illkirch Cedex

, France

Tel: 33 3 88 65 33 60

Fax: 33 3 88 65 32 01

Email: thisse@ibmc.u-strasbg.fr

EST from a cDNA of a gene whose expression is spatially restricted

during embryogenesis. We have established its expression pattern

during embryonic development by whole mount in situ hybridization

on zebrafish embryos from the gastrula stage to 2 days of

development. The corresponding data are available on the zebrafish

community database at <http://zfinfo.org/cDNA> library preparation: B.

Riggleman. DNA Sequencing by: IGBMC sequencing facility. Clone

distribution: zebrafish international resource center at the

University of Oregon (Institute of Neuroscience, 1254 University of

Oregon, Eugene, OR 97403-1254)

Seq primer: T3 ATTAACCTCTCAATAAGGA.

FEATURES

Location/Qualifiers

1..564

/organism="Danio rerio"

/mol_type="mRNA"

/db_xref="taxon:7955"

/clone="CB380"

/dev_stage="gastrula stage embryos"

/clone_lib="gastrula stage cDNA library"

/note="Vector: Lambda Zap; Site_1: EcoRI; Site_2: XhoI;

Oligo dT cDNA library constructed from RNA pooled from

gastrula stage zebrafish embryos"

BASE COUNT 148 a 133 c 155 g 128 t

ORIGIN

Alignment Scores:

Pred. No.: 507 Length: 564
Score: 41.00 Matches: 7
Percent Similarity: 100.00% Conservative: 4
Best Local Similarity: 63.64% Mismatches: 0
Query Match: 83.67% Indels: 0
DB: 14 Gaps: 0

US-09-965-594-26 (1-11) x CB417286 (1-564)

QY 1 GlySerValIleValGlyArgIleValLeu 11

Db 507 GGGCAGTGGTGTCTCATGGGAAGTGTACTG 539

RESULT 15

LOCUS

DEFINITION BI981053 584 bp mRNA linear EST 26-JUL-2002
tU38H04.x1 zebrafish adult brain Danio rerio cDNA clone
IMAGE:5332062 3' similar to TR:097894 097894 CCAAT/ENHANCER BINDING
PROTEIN BETA ;, mRNA sequence.

ACCESSION

BI981053

VERSION

BI981053.1

KEYWORDS

EST.

SOURCE

ORGANISM

Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes

; Cyprinidae; Danio.

REFERENCE

1 (bases 1 to 584)

AUTHORS

Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy

,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wyllie,T., Underwood

,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,

Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,

Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.

and Wilson,R.

WashU Zebrafish EST Project 1998

Unpublished

JOURNAL

COMMENT

Contact: Stephen L. Johnson

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: zbrafish@watson.wustl.edu

cDNA Library Preparation: John Ngai. cDNA Library Arrayed by:

Matthew Clark. DNA Sequencing by: Washington University Genome

Sequencing Center Clone distribution: Genome Systems, St. Louis,

Missouri (web address: www.genomesystems.com) (email contact:

info@genomesystems.com) and Research Genetics, Huntsville, Alabama

(web address: www.resgen.com) (email contact: info@resgen.com) and

Ressourcenzentrum Primatendatenbank, Berlin, Germany (web address:

www.rzpd.de)

Seq primer: -40UP

High quality sequence stop: 422.

FEATURES

Location/Qualifiers

1..584

/organism="Danio rerio"

/mol_type="mRNA"

/db_xref="taxon:7955"

/clone="IMAGE:5332062"

/sex="mixed male and female"

/tissue_type="brain"

/dev_stage="adult"

/lab_host="E. Coli DH10B"

/clone_lib="zebrafish adult brain"

/note="Vector: pZiPLOX; Site_1: NotI; Site_2: SalI;

Original library was constructed in lambdaZiPLOX. Mass

excision of the cDNA library was performed to yield

pZiPLOX plasmids. Insert check was done in original

library."

BASE COUNT 159 a 151 c 118 g 156 t

ORIGIN

Alignment Scores:
Pred. No.: 531 Length: 584
Score: 41.00 Matches: 7
Percent Similarity: 100.00% Conservative: 4
Best Local Similarity: 63.64% Mismatches: 0
Query Match: 83.67% Indels: 0
DB: 12 Gaps: 0

US-09-965-594-26 (1-11) x BI981053 (1-584)

QY 1 GlySerValValIleValGlyArgIleValLeu 11
DB 171 GGGCAGTGGTGTGTCATCGGAAGAGTTGTACTG 139

Search completed: August 31, 2003, 04:27:56
Job time : 112.667 secs